



STRING THEORY AND QUANTUM GRAVITY '91

Proceedings of the Trieste Spring School & Workshop
ICTP, Trieste, Italy April 15 – 26, 1991

Edited by

J. Harvey

R. Iengo

K. S. Narain

S. Randjbar-Daemi

H. Verlinde

World Scientific

STRING THEORY AND QUANTUM GRAVITY '91

This page is intentionally left blank



STRING THEORY AND QUANTUM GRAVITY '91

Proceedings of the Trieste Spring School & Workshop
ICTP, Trieste, Italy April 15 – 26, 1991

Edited by

J. Harvey *University of Chicago*

R. Iengo *ICTP/SISSA*

K. S. Narain *ICTP*

S. Randjbar-Daemi *ICTP*

H. Verlinde *Princeton University*



World Scientific

Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 9128

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 73 Lynton Mead, Totteridge, London N20 8DH

STRING THEORY AND QUANTUM GRAVITY '91

Copyright © 1992 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

ISBN 9789814538428

Printed in Singapore by General Printing Services Pte. Ltd.

PREFACE

As in previous years, the 1991 Trieste Spring School on String Theory and Quantum Gravity presented an overview of the most recent developments in the field. In the time since the previous school, there had been impressive progress in the understanding of two-dimensional quantum gravity and low-dimensional string models, and it was therefore no surprise that this subject was the main focus of the school. While on the one hand the further development of the discretized approach to random surfaces via matrix models had produced a wealth of explicit results, also the continuum theory based the Liouville model was revealing more and more of its secrets. Introductory lectures to both approaches were given by E. Martinec and L. Alvarez-Gaumé. More advanced topics and results in Liouville and non-critical string theory were explained in lectures by D. Kutasov, while I. Klebanov gave a clear exposition of the remarkably successful application of the matrix model techniques to string theory embedded in two space-time dimensions.

A second topic at the school were the properties of string theory in particular non-trivial backgrounds. Lectures by J. Harvey and C. Callan discussed the rich structure of string solitons and instantons, both from the target space and worldsheet point of view, and H. Verlinde showed how conformal field theory techniques can be used to describe the propagation of strings in the newly discovered two-dimensional black hole solution. A clear overview of recent developments in the study of W -algebras was given by A. Bilal. R. Dijkgraaf's lectures explained the structure of topological string theory and their relation to ordinary strings, while quantum groups were the topic of two lectures by L. Faddeev. Finally, this proceedings also contains the contribution from invited speakers at the 3-day Workshop, which was held immediately after the school.

This page is intentionally left blank

CONTENTS

Preface	v
Spring School on String Theory and Quantum Gravity:	
An Introduction to 2D Gravity and Solvable String Models <i>E. Martinec</i>	1
String Theory in Two Dimensions <i>I. R. Klebanov</i>	30
Some Properties of (Non)Critical Strings <i>D. Kutasov</i>	102
Topics in Liouville Theory <i>L. Alvarez-Gaumé and C. Gómez</i>	142
Black Holes and Strings in Two Dimensions <i>H. Verlinde</i>	178
Supersymmetric String Solitons <i>C. G. Callan, Jr., J. A. Harvey and A. Strominger</i>	208
Introduction to W-Algebras <i>A. Bilal</i>	245
Workshop on String Theory:	
Duality Invariant Effective String Actions and Automorphic Functions for (2,2) String Compactifications <i>D. Lüst</i>	281
Moduli Spaces and the Geometries of $N = 2$ SUSY <i>P. Fré and P. Soriani</i>	301
Transition from Critical to Non Critical String at High Temperature <i>C. Kounnas</i>	324

Canonical Quantization of $SL(2, \mathbb{R})$ Chern-Simons Topological Theories	338
<i>C. Imbimbo</i>	
Singular Vectors in Virasoro Verma Modules	355
<i>M. Bauer</i>	
Double Scaling limit in $O(N)$ Vector Models	367
<i>P. Di Vecchia</i>	
$O(N)$ Vector Field Theories in the Double Scaling Limit	371
<i>J. Zinn-Justin</i>	

STRING THEORY AND QUANTUM GRAVITY '91

This page is intentionally left blank

An Introduction to 2d Gravity and Solvable String Models

E. Martinec

Enrico Fermi Institute and Department of Physics*

University of Chicago, Chicago IL 60637

and

Dept. of Physics and Astronomy, Rutgers University

Piscataway, NJ 08854

1. Introduction

Two-dimensional gravity has undergone a thorough examination over the last few years, especially with the emergence of efficient calculational techniques stemming from the matrix model approach. Two dimensions is an arena where rather complex phenomena (confinement, chiral symmetry breaking, integrability, nonperturbative phenomena, solitons; the list is extensive) can be stripped of complications of higher dimensional kinematics and dynamics while hopefully retaining many of the physical features of the problem of interest. Thus we study two dimensional gravity as a model for exploring the structure and formalism of quantum gravity: the wavefunction of the universe, the Wheeler-DeWitt equation[1], the meaning of measurement in quantum gravity, the statistical properties of the metric and matter in a fluctuating geometry, etc. Indeed the collection of solvable 2d gravity-matter systems provides a rich laboratory for the investigation of these issues. In addition, since unified strings are by definition coordinate invariant 2d quantum field theories these systems are equally well regarded as solvable models of string theory, where one might begin to formulate a useful string field theory[2], study nonperturbative effects[3], strong coupling phenomena, etc.

There are several ways one might approach the continuum theory of 2d gravity. One is to write down a field theory on 2d metrics and matter[4], regulate it covariantly, and try to find a consistent renormalization to the continuum limit[5][6]. Another is to discretize the theory and study the fluctuations of the discrete geometry[7], then try to take the continuum limit of the random lattice model (a method that has also been used for strings embedded in higher dimensional spacetimes[8]). Both should yield the same results

* Permanent address

if indeed 2d gravity is universal, *i.e.* if there is a second order phase transition for the fluctuating surfaces in some region of the space of coupling constants (cosmological constant, topological coupling, boundary cosmological constant, etc.). The former line of attack adheres more closely to the standard conceptual framework of gravity, yet in two dimensions the latter formulation has yielded more success to date. This success comes about largely because in two dimensions it is not difficult to enumerate the lattice configurations, which are gauge invariant.

These lecture notes introduce the present understanding of 2d gravity/solvable string models, beginning in section 2 with a review of the continuum formulation in terms of the functional integral over metrics. While there is little that can be computed exactly, we argue that much of the classical structure should persist upon quantization; and that the familiar structures of rational conformal field theory might appear, albeit in a much more subtle and complicated way. In section 3 we discuss the behavior of the theory under rescalings of the 2d metric[9] as well as the related dynamics in the zero mode sector[5][10][11] [12]. In section 4 we switch to the discrete formulation. Here the partition function of random surfaces can be recast as an integral over $N \times N$ matrices[7] where in the continuum limit $N \rightarrow \infty$ [13]. Both the lattice[14] and continuum[13][15][16] theories exhibit a rich integrability structure. The scaling behavior predicted by the continuum approach is recovered as well as the zero mode dynamics[10][12].

2. The effective action for gravity and its quantisation

A crucial feature of 2d quantum gravity is that it must be scale invariant as well as reparametrization invariant[17]; local shifts $g_{ab} \rightarrow e^\epsilon g_{ab}$ of the metric scale factor are simply shifts of an integration variable in the functional integral over metrics. The effect of such a shift is to produce the trace of the stress tensor, which we can generally expand in terms of scaling fields

$$\langle T_{zz} \rangle = \frac{\delta}{\delta \epsilon(z)} \mathcal{Z} = \int e^{-S} \beta_i \mathcal{O}^i. \quad (2.1)$$

The β_i are the beta functions of the theory. The result (2.1) must vanish apart from possible contact terms with local operators. Often we are interested in gravity coupled to some number d of scalar matter fields

$$S = \frac{1}{4\pi} \int \sqrt{g} [T(X) + R^{(2)} D(X) + \partial_a X^\mu \partial_b X^\nu (G_{\mu\nu}(X) g^{ab} + B_{\mu\nu}(X) \epsilon^{ab})]. \quad (2.2)$$

in which case the vanishing trace of the stress tensor enforces a set of conditions on the spacetime fields which are the equations of motion of string theory. One may adopt either of two points of view: In ‘critical’ string theory, one sets $\frac{\delta}{\delta\epsilon}\mathcal{Z}_{matter} = 0$; then one has an additional symmetry – Weyl invariance – which one can quotient by, so that one integrates over the gravitational measure $\frac{\mathcal{D}g_{ab}\mathcal{D}X}{\text{Diff} \times \text{Weyl}}$. In this way one regards all components of the 2d metric as gauge degrees of freedom. On the other hand, from the ‘noncritical’ string point of view, one can trivially achieve scale invariance by integrating over all possible scale factors $\frac{\mathcal{D}g_{ab}\mathcal{D}X}{\text{Diff}}$. This is equivalent to a particular class of solutions to critical string theory if we regard the local scale factor as another scalar field like the X ’s[18]; perhaps it is completely equivalent if we complexify the space of metrics. In either case the object is to compute and solve the equations of vanishing stress tensor. There are two common methods of calculation: the spacetime weak field expansion[19] and the 2d loop expansion[20]. In the former one expands around a known fixed point (*e.g.* free field theory in d spacetime dimensions)

$$S = \frac{1}{4\pi} \int \partial X \bar{\partial} X + \delta T(X) + \partial X^\mu \bar{\partial} X^\nu (G_{\mu\nu} \delta D(X) + \delta G_{\mu\nu}(X) + \delta B_{\mu\nu}(X)) .$$

Bringing down powers of the composite operator perturbations will reliably find nearby fixed points (unless one has chosen a singular parametrization of the coupling space). The origin of the beta function is the overall scale divergence

$$\int \frac{d^2\lambda}{|\lambda|^2} \lambda^{L_0+L_0} \int_{N-1} \lambda^{L_0} \left\langle \prod_{i=1}^N \mathcal{O}_i(\lambda z) \right\rangle$$

The similarity of this expression to the Koba-Nielsen prescription for the calculation of the string S-matrix is not coincidental[21]. Indeed, the effective action of string theory (actually any field theory) is determined from the S-matrix by subtracting the contributions of intermediate on-shell poles. In the Koba-Nielsen formula these are the logarithmic subdivergences in the integrations over the locations of composite operators; the beta functions are the equations of motion following from this effective action. Note that the kinetic operator $\partial^2 S / \partial g_i \partial g_j = \partial \beta_i / \partial g_j$ for small fluctuations is the anomalous dimension operator – the linearized scaling operator $L_0 + \bar{L}_0 - 2$. The disadvantage of the spacetime weak field expansion is that one must work in a particular coordinate basis in 2d field space, so spacetime general coordinate invariance (invariance under 2d field redefinitions of the X ’s) is not manifest. This is an advantage of the 2d loop expansion; rather than

working in a specific basis of tensor fields on the spacetime manifold, one expands in small fluctuations of the coordinates X , which can be made manifestly generally covariant[20]. This perturbation series is an expansion in spacetime variation of the string coordinates relative to the 2d Planck constant, the inverse string tension α' . A superposition of these two methods yields the beta functions[20]

$$\begin{aligned}\beta_{\mu\nu}^G &= R_{\mu\nu} - 2\nabla_\mu \nabla_\nu D + \nabla_\mu T \nabla_\nu T = 0 \\ \beta^D &= \frac{26-d}{3\alpha'} + R + 4(\nabla D)^2 - 4\nabla^2 D + (\nabla T)^2 + V(T) = 0 \\ \beta^T &= -2\nabla^2 T + 4\nabla D \nabla T + V'(T) = 0\end{aligned}\tag{2.3}$$

in a double expansion in field strength and α' ; $V(T)$ is a generic potential $V(T) = \frac{1}{2}T^2 + O(T^3)$. Both expansions are required here; the loop expansion for general covariance, and the weak field expansion to incorporate the tachyon.

From the critical string viewpoint, we are interested in strings in d spacetime dimensions. The equations (2.3) are solved by ($\phi = X^0$, say)

$$\begin{aligned}T &= \frac{\mu}{2\gamma^2} e^{\gamma\phi} \quad , \quad Q = \frac{2}{\gamma} + \gamma = \sqrt{\frac{26-d}{3}} \\ D &= \frac{1}{\gamma} \phi \\ G_{\mu\nu} &= \delta_{\mu\nu} ;\end{aligned}\tag{2.4}$$

Although each equation of motion in (2.3) is satisfied at its leading nontrivial order in powers of T , none is solved at subleading order. However we are only considering the lowest order equations; one might hope that higher orders correct the problem. After all, the α' (loop) expansion is not valid here since γ is not small; the kinetic term in the tachyon beta function is only found after a resummation of loops, but then we have no reason to ignore terms involving the gradient of the dilaton field. Fortunately we are looking for a solution with rather special properties: it is (2.4) at lowest order and its renormalization involves only ϕ -dependent fields. Since $\nabla_\phi T \propto T$, the exact solutions $\hat{G}$, $\hat{D}$, $\hat{T}$ are power series in T (assuming the weak field expansion is summable). Therefore we can find a field redefinition – the reversion of the power series $\hat{G}(T)$, $\hat{D}(T)$, $\hat{T}(T)$ – such that (2.4) is the *exact* solution¹. The importance of $\nabla T \propto T$ is that the field redefinition required is local in spacetime.

¹ This argument is due to T. Banks.

The noncritical string viewpoint is somewhat more helpful in this situation. Here we have $d - 1$ string coordinates coupled to a dynamical metric $g_{ab} = e^{\gamma\phi} \hat{g}_{ab}$ where $\hat{g}_{ab}$ is a fixed metric (up to moduli). For $d - 1$ free scalar fields we have

$$\beta = \frac{26-d}{3\alpha'} \sqrt{g} R^{(2)} + \frac{\mu}{2\gamma} \sqrt{g}$$

Integrating $\delta S_{eff}/\delta\phi = \beta$ gives an effective classical action, the Liouville action

$$S_{eff} = \frac{1}{4\pi} \int \frac{1}{2} (\partial\phi)^2 + \frac{1}{\gamma} \phi R^{(2)} + \frac{\mu}{2\gamma^2} e^{\gamma\phi} \quad (2.5)$$

for the dynamics of the metric. The integral over metrics restores scale invariance if we can succeed in correctly performing the functional integral or otherwise quantizing the theory. From critical string considerations, one might expect a problem since the tachyon stress-energy appears to feed into the metric beta function. There is in fact a renormalization prescription such that S_{eff} is conformally invariant[17]. Since the metric is dynamical, the coordinate regularization scale and the covariant scale are related by a fluctuating field: $\delta = e^{\gamma\phi/2} \epsilon$. Therefore a regulated action should be, *e.g.*

$$S_{eff}^{reg} = \frac{1}{8\pi} \int d^2z \partial\phi (e^{\epsilon^2 e^{-\gamma\phi} \partial^2}) \bar{\partial}\phi + Q R^{(2)} \phi + \frac{\mu^2}{\gamma^2} e^{\gamma\phi},$$

and we should calculate $\delta S_{eff}/\delta\phi = \beta$ incorporating the effects of the regulator. In the loop expansion, normal ordering an exponential means resumming self-contractions; this introduces cutoff-dependence, and therefore ϕ -dependence. One finds that the entire effect of imposing $\beta_{eff} = 0$ is to renormalize the parameters in the Liouville lagrangian: $\mu \rightarrow \mu_{ren}$ (which can be absorbed in a shift of ϕ), and $Q \rightarrow Q_{ren} = \frac{2}{\gamma} + \gamma[5][17]$. Essentially we are allowed to “normal order” the Liouville term, and provided its quantum scale dimension is (1,1) and $\gamma \leq \frac{Q}{2}$ all is well. Physically, at short distances ($\phi \rightarrow -\infty$) the exponential potential term is small, so divergences should be those of free field theory. It seems that the complicated field redefinition of the critical string theory amounts to the change from coordinate to covariant regulator on the world sheet; recall that spacetime field redefinitions are equivalent to a change of renormalization scheme in the 2d sigma model. It is intriguing that in this low-dimensional context there can be nontrivial solutions of the string equations of motion whose stress-energy does not back-react on the spacetime metric.

One indication of the conformal invariance of quantum Liouville theory is the degree of singularity of the operator product expansion of the Liouville perturbation; assuming free field operator products (but see [11][10])

$$e^{\gamma\phi(z)} e^{\gamma\phi(w)} \sim |z - w|^{-2\gamma^2} e^{2\gamma\phi(z)} + \dots \quad (2.6)$$

In the weak coupling regime $\gamma \ll 1$ the singularity is integrable and the beta function calculated above will continue to vanish at first order in the perturbed theory. This reasoning does not give the correct upper bound on γ however, indicating that problems set in for $\gamma \geq 1$ ($c \geq -2$) rather than the observed upper bound $\gamma = \sqrt{2}$ ($c = 1$). Ordinarily in conformal field theory, an operator that produces the identity in its operator product with itself has

$$\mathcal{O}(z)\mathcal{O}(w) \sim |z-w|^{-4h}\mathbf{1}$$

from which one can read off the dimension of that operator. It seems to be that, even though it does not produce the identity, in its self-product the singularity of $e^{\gamma\phi}$ is four times its ‘effective dimension’. Then from (2.6) we would conclude that $e^{\gamma\phi}$ becomes effectively irrelevant at $c = 1$, or more generally $\mathcal{O}e^{\alpha\phi}$ becomes effectively irrelevant for $\alpha > Q/2$. This behavior is rather similar to what happens when attempting to incorporate massive string states in the sigma model lagrangian of the critical string. The on-shell vertex operator for such a perturbation has $h = 1$, but only by virtue of an irrelevant spatial operator coupled to a negative dimension temporal operator. The disease of irrelevant operators, namely nonrenormalizability, shows up in the appearance of many low dimension operators with highly singular coefficients in the self-OPE. Such a similar explosion occurs for operators $\mathcal{O}e^{\alpha\phi}$ with $\alpha > Q/2$ (for instance the ‘black hole perturbation’[22] $\partial X \bar{\partial} X e^{Q\phi}$, hence it is in no sense a ‘perturbation’ of flat spacetime Liouville theory).

To recapitulate, we have

$$\begin{aligned} T_{zz} &= 0 && \text{(Liouville e.o.m.)} \\ T_{zz} &= -\frac{1}{2}(\partial\phi)^2 + \frac{Q}{2}\partial^2\phi && Q = \frac{2}{\gamma} + \gamma \\ h_{\text{exp}[\alpha\phi]} &= -\frac{1}{2}\alpha(\alpha - Q) \end{aligned}$$

These look like free field results, but it must be stressed that $\langle\phi\phi\rangle$ is *not* the free field propagator. The Liouville equation is geometrically the equation for constant (negative for $\mu > 0$) curvature surfaces $R = -\mu$. The semiclassical expansion is valid for $\gamma \ll 1$; rescaling $\phi \rightarrow \frac{2}{\gamma}\phi$ puts an overall factor of $1/\gamma^2$ in front of (2.5). If we plot the potential for the zero mode

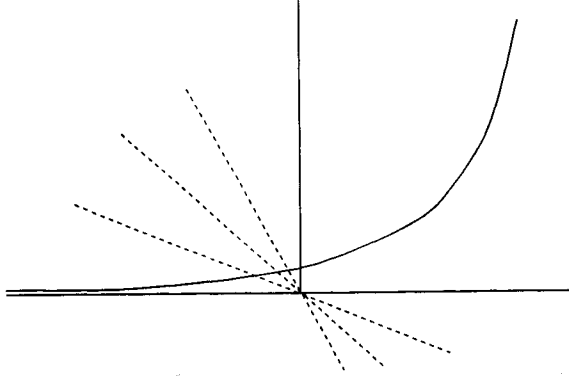


Fig. 1 Liouville potential (solid line) and curvature term (dashed line) for increasing genus.
The stable point of the total zero mode potential increases with the genus.

we see that a stable solution exists only for surfaces of genus $g > 1$ (in the absence of point sources of curvature; see below). The stable point $\langle \phi \rangle$ increases with genus; in fact there is a scaling relation[9]: let $\phi \rightarrow \phi + \frac{1}{\gamma} \log a$, then

$$\mathcal{Z}(g_{str}, \mu) = \mathcal{Z}(g_{str}, a\mu) a^{-(2-2g)\frac{\gamma}{2\gamma}},$$

implying

$$\mathcal{Z}_g(g_{str}, \mu) = C_g \mu^{(2-2g)\frac{\gamma}{2\gamma}}. \quad (2.7)$$

The full classical solution to the Liouville equation of motion is discovered through its geometrical role; $e^{\gamma\phi}$ must be a density under coordinate transformations, $e^{\gamma\phi(z)} = |\frac{\partial z'}{\partial z}|^2 e^{\gamma\phi(z')}$. Since the standard constant negative curvature metric on the upper half plane (UHP) is $ds^2 = dzd\bar{z}/(\text{Im } z)^2$, ϕ must look locally like

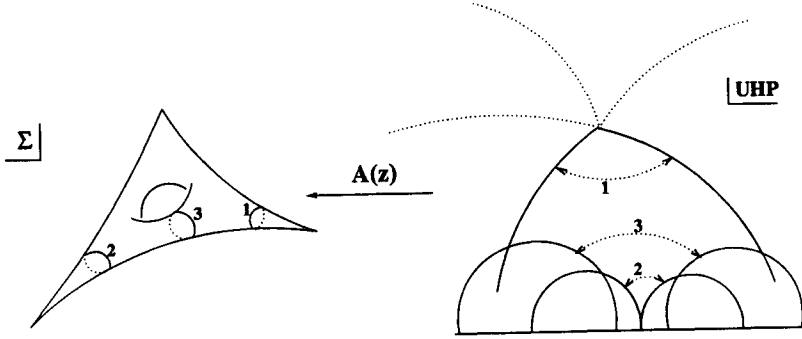
$$\phi = \frac{1}{\gamma} \log \left[\frac{16}{\mu} \frac{\partial A(z)}{(A - A^*)^2} \frac{\bar{\partial} A^*(\bar{z})}{(A - A^*)^2} \right],$$

where A, A^* are local coordinates on the surface; i.e. $A(z)$ is the map from the UHP to the Riemann surface Σ

$$\frac{\mu}{16} e^{\gamma\phi} dzd\bar{z} = \frac{dA dA^*}{(A - A^*)^2}.$$

The line element on the UHP is invariant under $\text{SL}(2, \mathbb{R})$ transformations $A \rightarrow \frac{aA+b}{cA+d} \equiv g(A)$. This transformation must leave ϕ invariant, but may do so in a nontrivial way, e.g.

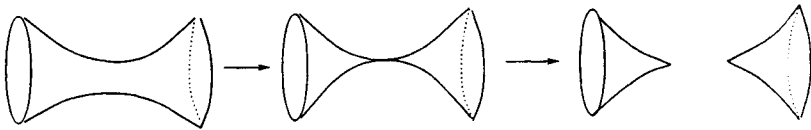
by making a circuit of a nontrivial closed path on Σ (see fig.2). That is, the *monodromy* of the local coordinate A is a set of $SL(2, \mathbb{R})$ transformations which cover the surface Σ onto the UHP. There are several classes of monodromies:



- (1) elliptic monodromy - $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is conjugate to a rotation, i.e. there exists h such that $hgh^{-1} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$. The surface Σ has a conical singularity of deficit angle θ . Deficit angles $\theta = \frac{2\pi}{j}$, $j \in \mathbb{Z}$, are 'nice' since they are covered by the UHP; $j = \frac{m}{n}$ requires an n -fold branched cover of the UHP at the fixed point of the rotation.
- (2) parabolic monodromy - h exists such that $hgh^{-1} = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}$, a translation (in terms of deficit angles, $\theta = \pi$ and the surface has a cusp).
- (3) hyperbolic monodromy - $hgh^{-1} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$ is a dilation. The identification of the UHP under g makes a handle.

A complete solution of the Liouville equation consists of representing the surface Σ by its fundamental group, the discrete subgroup (Fuchsian group) of $SL(2, \mathbb{R})$ that covers the surface onto the UHP; then one constructs the automorphic function $A(z)$ covariant under this group.

The three types of monodromy above are continuously related, as can be seen by pinching a handle on Σ :



In the classical theory, a deficit angle θ is a delta-function source of curvature:

$$\frac{1}{4\pi} \partial^2 \phi + \frac{\mu}{8\pi\gamma} e^{\gamma\phi} = \frac{Q}{8\pi} \hat{R} = \sum_i \frac{\theta_i}{\pi\gamma} \delta^{(2)}(z - z_i). \quad (2.8)$$

Integrating both sides, one sees that there is a solution whenever $2 - 2g + \theta_i/\gamma < 0$. Eq. (2.8) is the saddle point of the functional integral

$$\int \mathcal{D}\phi e^{-S_{\text{eff}}} \prod_i e^{\frac{\theta_i}{\pi\gamma} \phi(z_i)}. \quad (2.9)$$

On the other hand, there is no local source for hyperbolic monodromy (the fixed points of the $\text{SL}(2, \mathbb{R})$ transformation g on the UHP are not on the surface Σ). In terms of Liouville dynamics, the initial field configuration for ϕ (along some closed loop generating the monodromy) never propagates to $\phi = -\infty$ where the field configuration can be interpreted as localized at a point in the covariant metric. Note also that there is no classical geometrical interpretation for deficit angle $\theta > \pi$; *i.e.* when the source contributes half the curvature contribution $\frac{1}{8\pi} Q \int R$ of the sphere in the functional integral. Two parabolic cusps turn a sphere into a cylinder; we cannot go beyond this while maintaining the geometrical interpretation of ϕ . Quantum mechanically this means[10] $\alpha = \frac{\theta}{\pi\gamma} \leq \frac{Q}{2} = \frac{1}{\gamma} + \frac{\gamma}{2}$. This does not mean that operators with such Liouville charge ‘don’t exist’, rather merely that we cannot give them a geometrical interpretation. From the discussion of section (2), we conclude that gravitationally dressed operators with $\alpha > Q/2$ are ‘effectively irrelevant’. If we perturb the action by them, new dimension one operators will have to be added to the action to subtract singularities, a procedure that rapidly snowballs; as vertex operators they cannot be renormalized simply by normal ordering because the loop expansion is not well-behaved. It is doubtful that KPZ scaling (see below) can be maintained.

The above considerations motivate a brief review of $\text{SL}(2, \mathbb{R})$ representation theory[23]. Representations are labelled by their values of the quadratic Casimir $C_2 = j(j-1)$

and $J^3 = m + E_0$, $m \in \mathbb{Z}$. All representations can *formally* be built from the two-dimensional representation $\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \rightarrow \begin{pmatrix} g(w_1) \\ g(w_2) \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$. Realizing the $\text{SL}(2, \mathbb{R})$ algebra on differential operators $J^+ = \frac{1}{\sqrt{2}} w_1 \partial_2$, $J^- = \frac{1}{\sqrt{2}} w_2 \partial_1$, $J^3 = \frac{1}{2} (w_1 \partial_2 - w_2 \partial_1)$, the monomial $N_m w_1^a w_2^b = N_m (w_1 w_2)^j (w_1/w_2)^{E_0+m}$ transforms as part of the spin j representation, where N_m is a normalization. There are several cases:

- (1) The trivial representation $j = E_0 = m = 0$.
- (2) The finite dimensional representations $-2j \in \mathbb{N}$, $E_0 = 0$, $m = -j, \dots, j$; these are not unitary.
- (3) Discrete series representations $\mathcal{D}_\pm$: For $\mathcal{D}_+$, we have $-j + E_0 = 0$, $-2j \notin \mathbb{N}$, $J_3 + E_0 = 0, -1, -2, \dots$; for $\mathcal{D}_-$, $j + E_0 = 0$, $-2j \notin \mathbb{N}$, $J_3 + E_0 = 0, 1, 2, \dots$. These are unitary if $E_0 \in \mathbb{R}$ and $j > 0$. They are related to the elliptic monodromy conjugacy classes of $\text{SL}(2, \mathbb{R})$ and have $C_2 > 0$.
- (4) Continuous series representations: $-\frac{1}{2} < \text{Re } E_0 \leq \frac{1}{2}$, $-j + E_0 \neq 0, \pm 1, \pm 2, \dots$, $J_3 + E_0 = 0, \pm 1, \pm 2, \dots$. These are unitary if $E_0 \in \mathbb{R}$, $j - \frac{1}{2} \in i\mathbb{R}$; or $E_0, j \in \mathbb{R}$, $|-j + \frac{1}{2}| < \frac{1}{2} - |E_0|$. These are related to hyperbolic conjugacy classes in $\text{SL}(2, \mathbb{R})$ and have $C_2 < 0$.

How are these representations related to Liouville theory? Recall that classically $T_{zz} = -\frac{1}{2}(\partial\phi)^2 + \frac{1}{\gamma}\partial^2\phi$. From this follows [6][24]

$$(\partial^2 + \frac{\gamma^2}{2}T(z))e^{-\frac{\gamma}{2}\phi} = 0, \quad (2.10)$$

and from the classical solution we have

$$\begin{aligned} e^{-\frac{\gamma}{2}\phi_{cl}} &= \sqrt{\frac{16}{\mu}} \frac{(A - A^*)}{\sqrt{\partial A} \sqrt{\partial A^*}} \\ &= \sqrt{\frac{16}{\mu}} \sum_{m=1,2} \psi_{j,m}(z) \psi_{j,m}^*(\bar{z}) \end{aligned} \quad (2.11)$$

for $j = -\frac{1}{2}$, where

$$w_1 \equiv \psi_{-\frac{1}{2}, \frac{1}{2}} = \frac{1}{\sqrt{\partial A}} \quad ; \quad w_2 \equiv \psi_{-\frac{1}{2}, -\frac{1}{2}} = \frac{A}{\sqrt{\partial A}}.$$

are the two solutions to (2.10). The expressions (2.11) can be regarded as a classical version of conformal field theory where each measurement is a sum of holomorphic times antiholomorphic ‘chiral vertices’, glued together in a monodromy invariant way to make

physical single-valued expressions. From the two basic solutions one can in principal build all exponentials corresponding to finite-dimensional representations, *e.g.*

$$e^{j\gamma\phi} = \sqrt{\frac{16}{\mu}} \sum_{m=-j, \dots, j} \psi_{j,m}(z) \psi_{j,m}^*(\bar{z}) \quad , \quad -2j \in \mathbb{N}$$

An important issue is whether such expressions generalize to other classes of representations for which the sum over m is infinite. In principle one should be able to extract the answer from the classical solution $A(z)$ to (2.8), the saddle point of the correlation (2.9); this is in turn determined by its monodromy $\Gamma \in \text{SL}(2, \mathbb{R})$.

The quantum theory may have a structure similar to that of known conformal quantum field theories: the correlation functions would be (generically infinite) sums of holomorphic times antiholomorphic conformal blocks glued together to make a monodromy invariant object[25]

$$\langle \prod \mathcal{O}_i(z, \bar{z}) \rangle = \sum_{\alpha} \mathcal{F}_{\alpha}(z) \mathcal{F}_{\alpha}^*(\bar{z}) \quad . \quad (2.12)$$

For instance, in current algebra conformal field theories the conformal blocks $\mathcal{F}_{\alpha}$ carry two sets of indices, an ‘external’ index (like i in (2.12)) which is a representation label for the current algebra, and an ‘internal’ index which is a quantum group index, since the monodromy acts on the blocks like the R -matrix of a quantum group[26]. In the case of Liouville, there is no ‘external’ symmetry group other than Virasoro, and the index is continuous; however there is an ‘internal’ symmetry group $SL_q(2)$ which is, in some sense yet to be understood, the ‘quantization’ of the classical monodromy action on $A(z)$ [6][24]. In this respect Liouville is like other coset conformal field theories (Liouville can be thought of as the coset conformal field theory based on $SL(2, \mathbb{R})/N$, where N is the Borel subgroup[27]), where the external symmetry is gauged and the internal quantum group symmetry remains but is ‘confined’ – only appearing when one pulls the theory apart into holomorphic constituents. In terms of the parameter γ in the Liouville lagrangian, the quantum group parameter is $q = \exp[i\pi\gamma^2]$ [28][6]. The translation into $\text{SL}(2, \mathbb{R})$ notation of the condition $\alpha \leq Q/2$ means that $e^{j\gamma\phi}$ must have $j \leq \frac{Q}{2\gamma} = \frac{1}{\gamma^2} + \frac{1}{2}$; the quantum dimension is $h_j = -j - \frac{\gamma^2}{2}j(j-1)$. In these expressions the first term is the classical value; the second is the quantum correction. Note that L_0 is also real for $j = \frac{Q}{2\gamma} + i\lambda$, $\lambda \in \mathbb{R}$, for which $h_j = \frac{\gamma^2}{2}[(\frac{Q}{2\gamma})^2 + \lambda^2]$.

An analogy with the $\text{SU}(2)$ WZW theory might be helpful here. The classical solution to the theory is

$$g_{ab}(z, \bar{z}) = \sum_c g_{ac}(z) \bar{g}_{cb}(\bar{z})$$

where the holomorphic part $g(z)$ transforms under left multiplication by the loop group of $SU(2)$ and the Virasoro algebra, and also by right multiplication under $SU(2)$. The holomorphic constituents are determined by their monodromy $g(z) \rightarrow g(z) \cdot h$ around nontrivial closed loops on Σ . $g_{ab}(z)$ intertwines with $g_{cd}(z')$ via a classical r -matrix; upon quantization this internal index labels an $SU_q(2)$ quantum group representation with $q = \exp(\frac{2\pi i}{k+2})$ [26], and the intertwining is via a quantum R -matrix. For $SL(2)$ and for Liouville a major complication is that the fusion of representations closes on the continuous series representations and it is not clear that one has a useful or effective way of decomposing the correlations on intermediate states.

Although the above analysis is rather appealing as it places the Liouville theory within the customary conceptual framework of conformal field theory, it has not proven to be sufficiently powerful to enable the calculation of correlation functions. It is a useful route in the case of rational conformal field theories because the monodromy representations are finite, related to a closed system of differential equations that one may derive from various Ward identities[25]. In the case of Liouville theory, and also noncompact current algebra, the monodromy representations are infinite-dimensional; and the Ward identities are not sufficiently powerful to give a closed system of differential equations that will determine the correlation functions. Although finite dimensional representations $e^{-j\gamma\phi(z)}$ lead to finite order differential equations in z , general matter operators require infinite dimensional representations which don't satisfy any simple equations; such representations are required in intermediate states and therefore upon factorization, hence the dependence of correlations on the position of these operators is an open question. It seems that at least in the case of Liouville theory, other analytic techniques are available (see the lectures of D. Kutasov at this school, and references therein).

The above analysis points to a picture of 'quantum Riemann surfaces' where the classical solution $A(z)$ of Liouville theory is deformed into some kind of quantum conformal block, its classical $SL(2, \mathbb{R})$ monodromy deforming into an $SL_q(2)$ quantum group structure with q related to the coupling constant γ of Liouville theory.

3. Scaly behavior

Regardless of our understanding of how to compute Liouville correlation functions, there are several general statements we can make about those correlations *assuming* the

quantization preserves free-field ultraviolet behavior even in the presence of the exponential interaction term. The strongest ‘theoretical’ reason for such a presumption (made implicitly by the authors of [9] is that physical short distance as measured in the metric $ds^2 = e^{\gamma\phi} dz^2$ is where the Liouville potential is exponentially small, and so should not affect the free field results. The ‘experimental’ evidence for the validity of this assumption is the agreement of predicted scaling relations[9], and more recently tree-level S-matrix elements, with the continuum limit of discretized random surfaces in the matrix model[7].

What are the Liouville predictions? Suppose we wish to study conformal matter with central charge $c(=d)$ above coupled to 2d gravity. The matter theory will contain a set of scaling operators $\mathcal{O}_k^{matter}$ of scale dimension h_k . A generally covariant theory must make $\mathcal{O}_k^{matter}$ into a coordinate density of scale dimension $(h^{tot}, \bar{h}^{tot}) = (1, 1)$; for instance,

$$\mathcal{O}_k^{grav} = e^{\alpha_k \phi} \mathcal{O}_k^{matter} \quad (3.1)$$

with

$$h_{total} = -\frac{1}{2}\alpha_k(\alpha_k - Q) + h_k = 1. \quad (3.2)$$

Then the integrated correlation functions of $\mathcal{O}_k^{grav}$ are 2d general coordinate invariant. The exponential Liouville dependence of such operators will modify KPZ scaling in correlation functions, cf (2.9). The string theory interpretation of this *gravitational dressing* is that the Liouville field ϕ is the (Euclidean) time component of the string’s position in spacetime; the time (ϕ) dependence of (3.1) is just that of a linearized solution of the (Euclidean) spacetime equations of motion with energy $\alpha_k = j_k \gamma$; and the relation $-\frac{1}{2}(\alpha_k - Q)^2 + h_k = \frac{1-d}{24}$ between α_k and h_k is the mass shell condition for linear perturbations, i.e. h_k is the eigenvalue of the spatial Laplacian L_0^{matter} . This makes it clear that conformal matter is from the spacetime point of view the special class of stationary solutions of the string equations. From the previous analysis of Liouville theory, if $1 - h_k > \frac{Q^2}{8}$ then there does *not* exist a *local* generally covariant measurement corresponding to the matter operator $\mathcal{O}_k$; i.e. $\mathcal{O}_k^{matter}$ looks local in coordinates $(z, \bar{z})$, but its gravitational dressing requires hyperbolic monodromy $j_k = \frac{Q}{2\gamma} + i\lambda$ which has no local interpretation in the covariant theory. Seiberg[10] has dubbed such operators tachyonic since their spacetime mass shell condition (3.2) implies imaginary mass for the corresponding string state in spacetime. The reason that some perfectly sensible local operators may not have a local gravitational dressing is that each measurement perturbs the local geometry by making a small deficit angle α_k in the surface at the point of measurement; there are no probes ‘outside’ the

universe that can make such a measurement without perturbing the geometry. It can and does happen that when gravity is switched on the geometry is perturbed too violently by some operators to have a good, local continuum limit.

In conclusion, we can take away the following main lessons about Liouville theory coupled to conformal matter:

(1) KPZ scaling:

$$\mathcal{Z}_g = C_g \mu^{(2-2g)Q/2\gamma}$$

(2) Operator scaling dimensions $\mathcal{O}_k^{grav} = e^{\alpha_k \phi} \mathcal{O}_k^{matter}$ that shift the KPZ scaling relation by α_k in correlation functions. $-\frac{1}{2}(\alpha_k - \frac{Q}{2})^2 + h_k = \frac{1-d}{24}$.

(3) A ‘phase transition’ at $d = 1$ where the gravitationally dressed identity operator becomes tachyonic in the sense described above. There is no obvious reason why tree amplitudes might not be analytically continued as in $d = 26$ to yield a sensible classical string S-matrix, but (as at $d = 26$) loop amplitudes are infinite because tachyons cause infrared divergences in loop integrals even in Euclidean spacetime. As emphasized by Seiberg[10], this phase transition is not always at $c = 1$, but occurs whenever the spectrum has physical tachyons $h_{min} < \frac{d-1}{24}$.

Properties (1) and (2) are special to gravitationally dressed conformal matter; (3) is expected to be a generic feature persisting even when the matter theory is massive. One advantage of the matrix model, to which we turn next, is the ability to calculate the partition function and correlations even for massive matter.

The KPZ scaling relations suggest that a large part of Liouville dynamics is accounted for by the zero modes. Also $c < 1$ minimal models coupled to gravity have the tachyon as the only physical state (there are some additional physical states at nonstandard values of the ghost number[29] whose meaning is less clear); reparametrization invariance cancels the fluctuations of the longitudinal modes, leaving only the center of mass motion of the string when the spacetime is 1+1 dimensional. Indeed, it has been shown that free string propagation is accurately described by the quantum mechanics of the zero modes [12]² One can imagine solving the reparametrization invariance constraints $T_{00} = T_{01} = 0$ (or equivalently the BRST invariance constraints in conformal gauge) to show that physical states contain no excitations of the string’s nonzero modes; the remaining constraint is

² String interactions are not saturated by the zero modes, which has been interpreted in [12] as due to the violation of the single string physical state conditions by contact interactions in the vertices.

the Wheeler-deWitt equation $T_{00}^{zero\ mode} = 0$ on the zero modes. In a conformal matter theory coupled to gravity this separates into the dynamics of the matter zero modes, whose spectrum of scaling dimensions couples to the Liouville zero mode equation[5][10][11]

$$\left[-\left(\ell \frac{\partial}{\partial \ell}\right)^2 + 4\mu\ell^2 + \nu^2 \right] \Psi_0(\ell) = 0$$

with $\ell = e^{\gamma\phi/2}$ and $\nu = \frac{2}{\gamma}(\alpha - \frac{Q}{2})$. The appropriate solutions to this equation are the modified Bessel functions

$$\Psi_0(\ell) = (\nu \sin \pi\nu)^{1/2} K_\nu(2\sqrt{\mu}\ell) \quad (3.3)$$

Physically, the function $K_\nu = \frac{\pi}{2\pi \sin \nu\pi} [I_{-\nu} - I_\nu]$ is the linear combination of ‘incoming’ and ‘outgoing’ waves $I_{\pm\nu}(2\sqrt{\mu}\ell)$ which is exponentially damped in the infrared $\ell \rightarrow \infty$, indicating total reflection. Some of the strongest evidence for the viability of the conformal field theory approach to 2d gravity is the appearance of these wavefunctions within the matrix model.

4. The matrix model

The Hilbert space of Liouville is functionally infinite dimensional, however the subspace of physical states is much smaller – at most a countable number for $c \leq 1$. Naively each gauge invariance removes one canonical pair of variables, one by a choice of gauge and another by the gauge constraint (i.e. that the generator of gauge transformations annihilate the physical subspace – the analogue of Gauss’ law in QED). In 2d gravity there are two reparametrization degrees of freedom $\xi^a \rightarrow \tilde{\xi}^a(\xi)$. The time components of the metric are the Lagrange multipliers of the gauge constraints $T_{00} = T_{01} = 0$; the canonical degrees of freedom are the spatial metric $e^{\gamma\phi}$ and its conjugate momentum π_ϕ . Thus we have the possibility to gauge away up to one scalar matter fields’ worth of local degrees of freedom³. As usual in quantum gravity the difficulty in imposing the constraints lies in the Hamiltonian constraint $T_{00} = 0$, since this involves the way that the 2d worldsheet is carved up into spacelike hypersurfaces and therefore involves the (coordinate) time development. This makes it difficult to find nice global gauge invariant states. Nevertheless, the

³ Note that in this counting the Ising model has $d = \frac{1}{2}$ because $\pi_\psi \equiv \psi$ is ‘half’ a canonical pair.

lesson to be drawn is that it should pay to reformulate the path integral on the space of physical configurations because it is expected to be *much smaller* than the space of metrics plus matter field configurations. The simplest way to enumerate physical configurations is to discretize the 2d surfaces. Consider a 2d surface built by gluing together uniform squares each with sides of length ϵ (or triangles, pentagons, etc. – it turns out not to matter which[7] unless one artificially tunes couplings[30]).

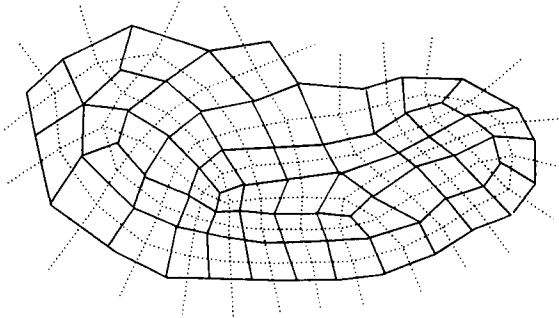


fig.4 Patch of discretized surface.

This lattice spacing replaces the covariant cutoff of Liouville theory. The only local freedom in pure gravity resides in how many squares meet at each vertex, which determines the local curvature discretized in units of $\pi/2$ (see fig.4). Note that each surface is dual to a Φ^4 graph (see fig.4), each square being dual to a Φ^4 vertex, each side of the square being dual to a ‘propagator’ of the Φ^4 Feynman graph. Thus counting all graphs with A vertices counts all surfaces with area $A\epsilon^2$; we can call this the partition function for discrete 2d Euclidean quantum gravity. The continuum limit consists of taking $\epsilon \rightarrow 0$, $A \rightarrow \infty$ such that the physical area $A_{phys} = A\epsilon^2$ is finite, assuming such a limit exists. In other words we concentrate on surfaces with a very large number of triangles; the statistics of these surfaces is governed by the large order asymptotics of graphical perturbation theory. We expect to be able to generate any local curvature in the continuum by coarse-graining over a large number of 3-, 4-, and 5⁺-coordinated vertices on the dual tessellation, of positive, zero and negative curvature, respectively.

In the pure gravity case, the generating function for the Φ^4 graphs dual to the discretization is the integral

$$\int d\Phi e^{-\frac{1}{2}\Phi^2 + g\Phi^4}, \quad (4.1)$$

i.e. the coefficient of g^A in the asymptotic expansion at small g is the number of surfaces with area A in lattice units, since this asymptotic expansion is the enumeration of Φ^4 Feynman graphs. The coupling g is to be identified with the bare 2d cosmological constant $g = \exp[-\mu_{bare}]$. Note that g is positive in order that each surface is counted with positive weight, so the generating function (4.1) diverges badly; we can interpret this through the asymptotic expansion as the statement that the entropy of large surfaces diverges uncontrollably. To cut down on this entropy we can try to count only surfaces of fixed genus, hoping that although the sum over topology is infinite, the individual terms in the series might be finite. This turns out to be the case for $d \leq 1$ matter.

Matter can be incorporated by introducing a label set for the dummy variable Φ [31]; then configurations are enumerated not only by the connectivity of the graph but also the element of the label set (which we think of as the value(s) of scalar field(s) or of discrete spin variables) at each point on the graph. For instance in the Ising model one considers two matrices M, N with integrand

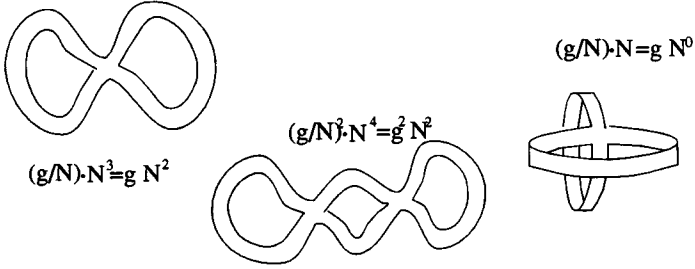
$$\exp(-\text{tr}[abM^2 + \frac{a}{b}N^2 + cMN + gdM^4 + \frac{g}{d}N^4]) .$$

The couplings a, b can be set to unity by a rescaling of M, N (however such redundant couplings can have physical effects through contact terms in loop correlations[32]). That leaves c , related to the Ising temperature because it controls the probability of transitions between up-spin (M) and down-spin (N) subgraph domains in the diagrammatic expansion; d is related to the magnetic field since it preferentially weights one of the two spins; and g is our friend the surface cosmological constant. The trick in this and similar cases is to take the continuum limit cleverly so that both graph connectivity and label fluctuations approach criticality. Fine-tuning more complicated potentials yields a tower of critical points[33].

To count the surfaces of fixed genus consider[34] the $N \times N$ Hermitian matrix field Φ_{ab} , and generating function

$$\int d\Phi e^{-\text{tr}[\frac{1}{2}\Phi^2 - (g/N)\Phi^4]} .$$

The reason to make Φ Hermitian is that the Feynman graphs carry an orientation and so the partition function counts orientable surfaces only. For the generalization to unoriented surfaces, see [35]. Looking at simple graphs



shows us that those which can be laid out smoothly on a surface of genus g come in the generating function with a factor N^{2-2g} . Each closed index loop traces over the indices in that loop and gives a factor N ; the coupling constant of the generating function is chosen to be g/N so that adding a square without changing the topology keeps the N counting fixed (the extra $\frac{1}{N}$ cancels the additional index trace). We have

$$\mathcal{Z}_{sft} = \int d\Phi e^{-N \text{tr} [\frac{1}{2} (\Phi/\sqrt{N})^2 - g (\Phi/\sqrt{N})^4]} = \exp \left[- \sum_g N^{2-2g} C_g \right]. \quad (4.2)$$

We have the right to call this generating function the partition function of (discretized) string field theory in this low-dimensional situation since it is indeed the object whose asymptotic expansion in N is the sum over surfaces of some 2d field theory describing the string background. The limit $N \rightarrow \infty$ picks out the sphere contribution C_0 , the classical limit of the associated string theory; we can evaluate this from the integral (4.2) by saddle point techniques even though the full integral doesn't make sense (the saddle is not a global minimum) since the leading contribution is just the value of the integrand at the saddle. To find the saddle, decompose the matrix in terms of 'matrix polar coordinates' as

$$\Phi = U \Lambda U^{-1}$$

where U is a unitary matrix and Λ is the diagonal matrix of eigenvalues of Φ . The partition function becomes

$$\mathcal{Z}_{sft} = \int U^{-1} dU \int d\Lambda \left| \frac{\partial \Phi}{\partial (U, \Lambda)} \right| \exp \left[-N \sum_{i=1}^N \left(\left(\frac{\lambda_i}{\sqrt{N}} \right)^2 - g \left(\frac{\lambda_i}{\sqrt{N}} \right)^4 \right) \right].$$

The Jacobian is easily evaluated by noting that a) it vanishes whenever any two eigenvalues coincide; b) it is symmetric under permutations of the eigenvalues; and c) it must scale like

$\lambda^{N(N-1)}$; the unique function with these properties is $\prod_{i < j} (\lambda_i - \lambda_j)^2$. The justification for a) is analogous to the vanishing of the Jacobian from Cartesian to polar coordinates in ordinary multidimensional integrals: the origin is a fixed point of the symmetry group of rotations. The transformation is singular there so the measure must vanish. Similarly, the coincidence of two eigenvalues is a fixed point for an $SU(2)$ subgroup of $U(N)$, so the Jacobian is singular. The partition function becomes

$$\mathcal{Z}_{sft} = \int \prod d\lambda e^{-N \sum_i V(\lambda_i/\sqrt{N}) + 2 \sum_{i < j} \log|(\lambda_i - \lambda_j)/\sqrt{N}|}.$$

The mental picture to adopt is to consider the eigenvalues as a set of particles lying in a metastable well, interacting through a logarithmically repulsive ‘Coulomb’ force. As $N \rightarrow \infty$ we can replace the set $\{\lambda_i\}$ by an eigenvalue density (which we might regard as a collective coordinate or string field[2]) and the saddle point equation is

$$\frac{1}{2}\lambda - 2g\lambda^3 = \int_{-a}^a \frac{\rho(\mu)d\mu}{\lambda - \mu}, \quad \int_{-a}^a \rho(\mu)d\mu = 1.$$

Consideration of analytic properties provides the solution[34]

$$\rho(\lambda) = \frac{1}{\pi} \left[\frac{1}{2} - ga^2 - 2g\lambda^2 \right] \sqrt{a^2 - \lambda^2} \quad \lambda \leq a = \left[\frac{-1}{6g} [(1 - 48g)^{1/2} - 1]^{1/2} \right].$$

Near the endpoint a of the distribution, $\rho(\lambda)$ behaves as $\rho(\lambda) \sim \sqrt{a^2 - \lambda^2}$; but when $g \rightarrow g_c = -\frac{1}{48}$ the analytic behavior changes to $\rho(\lambda) \sim (a^2 - \lambda^2)^{3/2}$. What is happening? The edge of the eigenvalue density becomes softer because the last eigenvalue approaches an instability where the Coulomb repulsion of the rest of the eigenvalues overcomes the external potential force and pushes it out of the metastable well. This instability reflects a phase transition in the surface dynamics on the sphere: recall that the entropy of surface configurations grows with the number of plaquettes, but there is an energy cost $\exp[-\mu A]$. At $g \sim g_c = e^{-\mu_c}$ from above these balance, and $\mathcal{Z}$ is dominated by surfaces controllably large compared to the cutoff ϵ . Inserting $\rho(\lambda)$ into the saddle point action gives

$$\begin{aligned} \frac{1}{N^2} S_{saddle} &= \int_{-a}^a \rho(\lambda) V(\lambda) - \int \int \rho(\mu) \rho(\lambda) \log|\lambda - \mu| \\ &= \frac{1}{24} (a^2 - 1)(9 - a^2) - \frac{1}{2} \log a^2 \\ &\sim C_0 (g - g_c)^{5/2} \end{aligned}$$

where $\epsilon^2 \mu_{ren} = g - g_c$. This saddle point action is the partition function of 2d gravity on the sphere, $S_{saddle} = Z_{sphere}^{2d} \sim C_0 N^2 (\epsilon^2 \mu_{ren})^{5/2}$. Letting

$$g_{str} = \frac{1}{N^2 \epsilon^{5/2}}$$

gives $Z_{sphere} = C_0 (\mu_{ren}^{5/2} / g_{str}^2) \equiv C_0 \mu^{5/2}$; in general the full string partition function depends on the couplings μ_{ren} and g_{str} only through the ratio μ . Note that to get a sensible continuum result we must let N scale with the surface cutoff $\delta \sim e^{\gamma\phi/2}$, i.e. g_{str} is dynamical – just what KPZ scaling predicts! Plugging numbers into the Liouville formulae (2.4),(2.7) we find $Q = 5\sqrt{3}$, $\gamma = 2/\sqrt{3}$ for $d = 0$, hence Liouville theory predicts

$$Z_{sphere}^{Liouv} = C_0 \mu^{\frac{Q}{2\gamma} \cdot 2} = C_0 \mu^{5/2}$$

as observed!

Note that the important feature of the phase transition used to take the continuum limit is the quadratic maximum in the effective eigenvalue potential (linear vanishing of the force on the last eigenvalue at $g = g_c$), so it shouldn't matter whether we use Φ^4 squares or Φ^3 triangles in the microscopic theory. In other words, only the rate of vanishing of ρ near its endpoint is universal; the details of the eigenvalue distribution far from this region as well as the value of g_c are irrelevant quantities. By fine tuning the relative weight of different polygonal simplices one can however reach other sorts of critical points[30][36].

The advance embodied in the discrete approach is the ease with which one may calculate the coefficients C_g in the partition function[13], as well as all correlation functions. To date it has only been possible in the Liouville approach to calculate (after much effort) correlation functions on the sphere. To go beyond the sphere one needs to among other things incorporate corrections to $\rho(\lambda)$ due to the discreteness of eigenvalues; ρ is not a smooth function at the $1/N$ level but rather a sum of delta functions. In fact the whole methodology above is rather cumbersome, and it proves simpler to reformulate the problem. It is apparent that the main difficulty is the Coulomb interaction among eigenvalues, so we might try to find variables that diagonalize this interaction as much as possible. This happens[34] when one writes the square root of the Jacobian as a Slater determinant

$$\prod_{i < j} (\lambda_i - \lambda_j) = \det_{ij} [\psi_i(\lambda_j)] \equiv |\Psi(\lambda)\rangle \ ,$$

where $\psi_i(\lambda) = \lambda^i - 1 + c_{i-2}\lambda^{i-2} + \dots + c_0$. The c_i are arbitrary since they correspond to $\psi_i \rightarrow \psi_i + \psi_k$, $k < i$, which cancels from the determinant. A convenient choice is to orthogonalize ψ with respect to the measure $e^{-V(\lambda)} d\lambda$

$$\mathcal{Z}_{sft} = \int \prod d\lambda \left\langle \Psi_t(\vec{\lambda}) \right| e^{-\sum_i V(\lambda_i)} \left| \Psi_t(\vec{\lambda}) \right\rangle \equiv \left\langle \Psi_t(\vec{\lambda}) \right| \Psi_t(\vec{\lambda}) \rangle$$

$$V(\lambda) = \sum_k t_k \lambda^k$$

Letting

$$h_n \delta_{nm} = \int d\lambda \psi_n(\lambda) e^{-V(\lambda)} \psi_m(\lambda), \quad (4.3)$$

we find

$$\mathcal{Z}_{sft} = \prod_{i=1}^N h_i$$

up to a t -independent normalization. The computation of the partition function reduces to the determination of the $h_i(t)$. To this end, define

$$h_n Q_{nm} = \int \psi_n(\lambda) e^{-V(\lambda)} \lambda \psi_m(\lambda) \quad (4.4)$$

$$h_n P_{nm} = \int \psi_n(\lambda) e^{-V(\lambda)} \frac{d}{d\lambda} \psi_m(\lambda). \quad (4.5)$$

Considering different matrix elements yields

$$\mathbf{Q} = \begin{pmatrix} b_1 & a_1 & & & 0 \\ 1 & b_2 & a_2 & & \\ & 1 & b_3 & a_3 & \\ 0 & & \ddots & \ddots & \ddots \end{pmatrix}. \quad (4.6)$$

This also tells us (by acting λ to the left) that $a_i = h_{i+1}/h_i$; a little thought shows $b_i = 0$ if V is an even function (since then we have the symmetry $\psi_j(\lambda) = (-)^j \psi(-\lambda)$). Integration by parts for $\mathbf{P}$ gives

$$\mathbf{P} = V'(\mathbf{Q})_+, \quad (4.7)$$

where the $+$ subscript means taking only the upper triangular part, setting the lower triangular and diagonal entries to zero, since $d/d\lambda$ must lower the order of ψ ; similarly a $-$ subscript refers to the lower triangular part. Now

$$[\frac{d}{d\lambda}, \lambda] = 1 \implies [\mathbf{P}, \mathbf{Q}] = \mathbf{1} \quad (4.8)$$

is an equation for a, b which solves our problem, *i.e.*

$$\mathcal{Z}_{sft} = \prod_{i=1}^N h_i = h_0^N \exp\left[\sum_{i=1}^N (N-i) \log a_i\right].$$

Afficionados of integrable systems will recognize that if $V = \frac{1}{2}\lambda^2$ then $\mathbf{P}, \mathbf{Q}$ is the Lax pair of the Toda chain; actually $\text{tr}\{Q^{k+1}\}$ is the k^{th} conserved charge of the Toda problem and $(Q^k)_+$ is the k^{th} Toda Hamiltonian. In other words,

$$\frac{\partial \mathbf{Q}}{\partial t_k} = [\mathbf{Q}_+^k, \mathbf{Q}] \quad (4.9)$$

are commuting flows. One easily checks that these flows preserve the string equation (4.8). Finally one can show that $\mathcal{Z}_{sft}$ is a particular type of ‘tau-function’ of the Toda system; this conveys little information but sounds impressive, the content being primarily in equations (4.8) and (4.9). Commuting flows in this problem should not surprise us; we must have $\frac{\partial^2 \mathcal{Z}}{\partial t_i \partial t_j} = \frac{\partial^2 \mathcal{Z}}{\partial t_j \partial t_i}$, etc., so of course all the derivatives commute. The nontrivial fact is that we can represent $\mathcal{Z}$ as an inner product in a Hilbert space

$$\mathcal{Z}_{sft}(t) = \langle \Psi_{\mathbf{t}} | e^{-\sum (t_k - i_k) \hat{Q}^k} | \Psi_{\mathbf{t}} \rangle \quad (4.10)$$

where the derivatives with respect to the couplings act as operators. The different flows in coupling space then form an integrable system in this Hilbert space.

Again we need to take the continuum limit. To this end it is convenient to absorb a factor $e^{-V/2}$ in ψ , normalize $\psi_k \rightarrow \frac{1}{\sqrt{h_k}} \psi_k$ and take V even (although the latter is not essential). Then the Lax pair $\mathbf{P}, \mathbf{Q}$ are transformed into

$$\tilde{\mathbf{Q}} = \mathbf{Q}_+ - \mathbf{Q}_- \quad \tilde{\mathbf{P}} = V'(\mathbf{Q})_+ - V'(\mathbf{Q})_-.$$

The scaling limit consists of taking $a_i \rightarrow \text{const.} + \epsilon^2 u(x)$, where $\frac{i}{N} = 1 - \epsilon^2 x$. The second order difference operator $\tilde{\mathbf{Q}}$ becomes (in the sense of matrix elements) a second order differential operator $\partial_x^2 + u(x)$; $a_N \approx \frac{\mathcal{Z}_{N+1} \mathcal{Z}_{N-1}}{\mathcal{Z}_N^2}$ yields $u(x) = \partial_x^2 \mathcal{F}$ where $\mathcal{Z}_{sft} = e^{-\mathcal{F}}$. Finally $\tilde{\mathbf{P}}$ scales to some odd order differential operator depending on the details of the potential (we made the rescaling in the wavefunctions in order to make this manifest). Generically $\tilde{\mathbf{P}} \rightarrow \partial_x$ which is boring, *i.e.*

$$[\tilde{\mathbf{P}}, \tilde{\mathbf{Q}}] = \mathbf{1} = u' = \mathcal{F}''' \implies \mathcal{F} = x^3$$

which corresponds to $\rho \sim \sqrt{a^2 - \lambda^2}$. Fine tuning as before should yield $\rho \sim (a^2 - \lambda^2)^{3/2}$ corresponding to $\tilde{\mathbf{P}} \rightarrow \partial^3 + v_1(x)\partial + v_0$. Then the equation $[\tilde{\mathbf{P}}, \tilde{\mathbf{Q}}] = 1$ is the continuum statement

$$u''' + uu' = 1 ,$$

the Painlevé I equation. This equation embodies through its asymptotic expansion in $x = \mu$ the entire perturbation series for $\mathcal{Z}_{sf t}$ about this background. The leading solution at large x is $u^2 = x$, which gives $\mathcal{F} \sim x^{5/2}$ which is the famous KPZ scaling relation! Plugging this solution into the differential equation and iterating yields

$$\mathcal{F} = \sum_{g=0}^{\infty} C_g x^{\frac{5}{4}(2-2g)}$$

so we see $Q/\gamma = 5/2$. Note that derivative terms in the string equation come with factors of the string loop coupling $g_{str} \sim 1/N$, which is why the genus expansion expands in powers of derivatives of $u(x)$. Douglas[16][37] has generalized this construction for an arbitrary pair of differential operators $\mathbf{P}, \mathbf{Q}$ of integer order p, q satisfying (4.8), which produces BPZ minimal matter[25] coupled to 2d gravity. Changing the bare matrix potential scales to continuum perturbations of P, Q by lower or higher order differential operators for relevant and irrelevant perturbations, respectively[37][38]⁴. In the two (and more) matrix case, we can again eliminate the angle variables in the partition function[34]. To ‘diagonalize’ the eigenvalue interaction, choose independently the left and right wavefunctions $\langle \Psi_t |$ and $|\Psi_t\rangle$ [39]. Defining matrices M, P_M and N, P_N analogous to (4.4) and (4.5), we have the variational equations

$$P_M = N + V'_M(M)$$

$$P_N = M + V'_N(N)$$

The number of nonzero diagonals of M, N are determined by the degrees of V_N, V_M respectively. Tuning these potentials, we can make M, P_M scale to a pair of differential operators

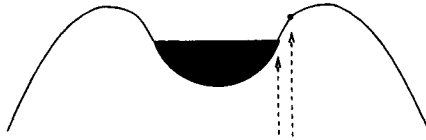
$$Q = \partial^q + u_{q-2}\partial^{q-2} + \dots + u_0$$

$$P = \partial^p + v_{p-1}\partial^{p-1} + \dots + v_0$$

of orders (p, q) . The Heisenberg relation $[P_M, M] = 1$ again determines the coefficients $u_n(x, t)$, $v_m(x, t)$ and hence the free energy $\partial_x^2 \mathcal{F} = u_{q-2}$ [16][37]. These critical points

⁴ Of course any fixed lattice operator is a sum over continuum operators with cutoff dependent coefficients.

have been identified with (p, q) conformal matter coupled to 2d gravity. The flows in the couplings t are governed by the 2d Toda hierarchy on the lattice[14], which scales to the q -reduced KP hierarchy in the continuum[16][37][38]. The coefficients $C_g \sim (2h)!$ for large g , reflecting the divergence of the sum of the perturbation series[13][13]. It is very interesting that these coefficients can be associated with an auxilliary, higher-action saddle of the matrix integral[3] corresponding to the eigenvalue configuration where the last eigenvalue is moved to its unstable equilibrium



which might be interpreted as a ‘string instanton’ mediating some kind of vacuum decay, although the precise meaning of this configuration remains unclear.

An interesting and calculable set of correlation functions in these systems are correlations of the loop operator $W(\ell)$ which cuts a hole of boundary length ℓ in the surface. In the matrix model this operator is $\frac{1}{l} \text{tr} \Phi^l$, which in graphs inserts a source of l external lines graphically dual to a loop or hole of boundary length $l\epsilon$. For instance, fig.4 is a contribution to $l = 24$. In the continuum limit $Q \rightarrow \text{const.} + \epsilon(-\partial^2 + u)$, $l \rightarrow \ell/\epsilon$ this becomes the heat operator $e^{-\ell(-\partial^2 + U)}$

$$\text{tr} \Phi^l \sim (\text{const.} + \epsilon(-\partial^2 + u))^{\ell/\epsilon} \sim e^{-\ell(-\partial^2 + u)} \equiv W(\ell)$$

up to an unimportant (nonuniversal) normalization. The calculation of loop correlation functions thus reduces to a set of convolutions of heat kernels[15], *e.g.*

$$\langle W(\ell_1) W(\ell_2) \rangle = \int_{-\mu}^{\infty} dx \int_{-\infty}^{-\mu} dy \langle x | e^{-\ell_1 Q} | y \rangle \langle y | e^{-\ell_2 Q} | x \rangle . \quad (4.11)$$

For the general (p, q) model we simply use the operator Q determined by the string equations (4.8). The loop correlations serve as a kind of generating function of local operator

correlations, in the sense that the $\ell \rightarrow 0$ limit of $W(\ell)$ can be recast as a sum (integral) over the spectrum of local operators[12]. Geometrically, the zero mode $\ell = \oint e^{\gamma\phi/2}$ is dual (in the sense of Fourier transform) to its conjugate Liouville momentum $p = \oint \pi_\phi$ whose eigenvalue is the deficit angle; hence it should be no surprise that a Green's function at fixed small ℓ should be an integral transform over local curvatures. The small ℓ expansion of (4.11) is the asymptotic expansion of the heat kernel, which is evaluated in terms of formal fractional powers of Q [40][15][12]. For instance, (4.11) expands as

$$\begin{aligned} \langle W(\ell)W(\ell') \rangle &\stackrel{\ell'\rightarrow 0}{\sim} \sum_{n=0}^{\infty} \ell'^{n/q} \langle \mathcal{O}_n^{KP} W(\ell) \rangle \\ \langle \mathcal{O}_n^{KP} W(\ell) \rangle &= \int^\mu dx [Q_+^{n/q}, \langle x | e^{-\ell Q} | x \rangle] . \end{aligned}$$

For the definition of fractional powers of differential operators and their uses, see *e.g.* [38][37][41]. The basis of operators defined by isolating individual terms on the RHS do not coincide with the dressed operators $\mathcal{O}_\alpha$ of conformal field theory[12]. Instead, one must take linear combinations of $\mathcal{O}_n^{KP}$ with coefficients defined by the ‘incoming wave’ $I_{n/q}(2\sqrt{\mu}\ell)$. Then one finds the two-point function on the sphere of a CFT scaling operator and a loop operator is

$$\langle \mathcal{O}_\alpha W(\ell) \rangle = (2\sqrt{\mu})^\alpha K_\alpha(2\sqrt{\mu}\ell) . \quad (4.12)$$

Thus one can justifiably claim that the matrix and Liouville approaches coincide where calculations can be done in both. Moreover, according to the ideas of Hartle and Hawking[1], one expects that the correlation function of a scaling operator $\mathcal{O}_k$ with a loop operator $W(\ell)$ is the corresponding wavefunction of that operator – the operator creates the appropriate state in the Hilbert space of physical states, and $W(\ell)$ is the probability of that state propagating on a Euclidean surface of average curvature $-\mu$ to a boundary of length ℓ . This expectation is indeed borne out by the explicit calculations [10][12] outlined above.

The matrix model is an effective tool for the calculation of correlations of integrated local scaling operators. While this is an important class of measurements it by no means exhausts the list of interesting questions one can try to ask in 2d gravity. These integrated correlations are what might be called extensive measurements, things like the area of the surface, or the magnetization or average energy density in the Ising model. Truly local measurements, where one tries to ‘build a laboratory’ on a patch of the surface and make measurements in it, have not been addressed yet because they are intrinsically more complicated. One must describe the laboratory in a coordinate invariant way, *e.g.* by

prescribing the geodesic distances and relative orientation of the objects in it. Even the simplest such ‘local’ measurement, the Hausdorff dimension of the surface, turns out to be very difficult to investigate[42]. It is of course possible that, like the scaling operators of large positive Liouville momentum, such measurements in two dimensions are plagued by strong fluctuations that render local questions meaningless. In such a situation the simple model does not retain the flavor of its higher dimensional cousin. Another interesting issue is whether we can define Minkowski 2d gravity independent of working in Minkowski string theory. Usually we induce the continuation to Minkowski signature world sheet through the analytic continuation of string amplitudes from Euclidean spacetime and its associated $i\epsilon$ prescription. Nevertheless it would be odd if the beautiful physics of $1+1$ -dimensional scattering theory could not be put on a fluctuating geometry. Can we not discuss the S-matrix of the Ising model, or Potts or sine-gordon models, in deSitter 2d gravity? This also may hinge on how to define localized asymptotic states in a fluctuating background geometry.

References

- [1] J. Hartle and S. Hawking, *Phys. Rev.* **D28** (1983) 2960; B.S. DeWitt, *Phys. Rev.* **160** (1967) 1113.
- [2] S. Das and A. Jevicki, *Mod. Phys. Lett.* **A5** (1990) 1639.
- [3] S. Shenker, Rutgers preprint RU-90-47, to appear in the proceedings of the Cargèse workshop (1990).
- [4] A.M. Poyakov, *Phys. Lett.* **103B** (1981) 207.
- [5] T. Curtright and C. Thorn, *Phys. Rev. Lett.* **48** (1982) 1309; *Ann. Phys.* **147** (1983) 365; *Ann. Phys.* **153** (1984) 147
- [6] J.-L. Gervais and A. Neveu, *Nucl. Phys.* **B199** (1982) 59; **B209** (1982) 125; **224** (1983) 329; **238** (1984) 125, 396; *Phys. Lett.* **151B** (1985) 271.;
J.-L. Gervais, *Phys. Lett.* **243B** (1990) 85; *Comm. Math. Phys.* **130** (1990) 257.
- [7] F. David, *Nucl. Phys.* **B257**[**FS14**] (1985) 433; V. Kazakov, *Phys. Lett.* **119A** (1986) 140; **150B** (1985) 282; V. Kazakov, I. Kostov, and A.A. Migdal, *Phys. Lett.* **B157** (1985) 295; J. Ambjorn, B. Durhuus, and J. Frohlich, *Nucl. Phys.* **B257** (1985) 433.
- [8] I. Klebanov and L. Susskind, *Nucl. Phys.* **B309** (1988) 175.
- [9] V. Knizhnik, A.M. Poyakov, and A.B. Zamolodchikov, *Mod. Phys. Lett.* **A3** (1988) 819; F. David, *Mod. Phys. Lett.* **A3** (1988) 1651; J. Distler and H. Kawai, *Nucl. Phys.* **B321** (1989) 509.
- [10] N. Seiberg, *Notes on Liouville Theory and Quantum Gravity*, *Prog. Theor. Phys., Suppl.* **102** (1990) 319.
- [11] J. Polchinski, 'Remarks on the Liouville field theory', Texas preprint UTTG-19-90, presented at Strings '90, College Station, 1990.
- [12] G. Moore, N. Seiberg, and M. Staudacher, *Nucl. Phys.* **362** (1991) 665; G. Moore and N. Seiberg, Rutgers preprint RU-91-29.
- [13] E. Brézin and V. Kazakov, *Phys. Lett.* **236B** (1990) 14; M. Douglas and S. Shenker, *Nucl. Phys.* **B335** (1990) 635; D. Gross and A.A. Migdal, *Phys. Rev. Lett.* **64** (1990) 127; *Nucl. Phys.* **340** (1990) 333.
- [14] L. Alvarez-Gaumé, C. Gomez, and J. Lacki, *Phys. Lett.* **253B** (1991) 56; A. Gerasimov, A. Marshakov, A. Mironov, A. Morozov, and A. Orlov, *Nucl. Phys.* **B357** (1991) 565; E. Martinec, *Comm. Math. Phys.* **138** (1991) 437.
- [15] T. Banks, M. Douglas, N. Seiberg, and S. Shenker, *Phys. Lett.* **238B** (1990) 279.
- [16] M. Douglas, *Phys. Lett.* **238B** (1990) 176.
- [17] D. Friedan, in the proceedings of the 1982 Les Houches Summer School; and unpublished.
- [18] F. David and E. Gitter, *Europhys. Lett.* **3** (1987) 1169; *Nucl. Phys.* **B293** (1988) 332; S.R. Das, S. Naik, and S. Wadia, *Mod. Phys. Lett.* **A4** (1989) 1033; J. Polchinski, *Nucl. Phys.* **B324** (1989) 123; T. Banks and J. Lykken, *Nucl. Phys.* **B331** (1990) 173.

- [19] C. Lovelace, *Nucl. Phys.* **B273** (1986) 413.
- [20] D. Friedan, *Phys. Rev. Lett.* **45** (1980) 1057; C. Callan, D. Friedan, E. Martinec, and M. Perry, *Nucl. Phys.* **B262** (1985) 593; E.S. Fradkin and A. Tseytlin, *Phys. Lett.* **158B** (1985) 316; *Nucl. Phys.* **B272** (1986) 647.
- [21] J. Hughes, J. Liu, and J. Polchinski, *Nucl. Phys.* **B316** (1989) 15.
- [22] E. Witten, IAS preprint IASSNS-HEP-91/12; G. Mandal, A. Sengupta, and S. Wadia, IAS preprint IASSNS-HEP-91/10.
- [23] see, e.g., A.O. Barut and C. Fronsdal, *Proc. Royal Soc.* **A287** (1965) 532.
- [24] P. Zograf and L. Takhtadjan, *Math. USSR Sbornik* **60** (1988) 143.
- [25] A. Belavin, A. Polyakov, and A.B. Zamolodchikov, *Nucl. Phys.* **B241** (1984) 333.
- [26] A. Alekseev and S. Shatashvili, *Comm. Math. Phys.* **133** (1990) 353.
- [27] A. Polyakov, *Mod. Phys. Lett.* **A2** (1987) 893; V. Kniznik, A. Polyakov, and A.B. Zamolodchikov, *Mod. Phys. Lett.* **A3** (1988) 819; A. Alekseev and S. Shatashvili, *Nucl. Phys.* **B323** (1989) 719; M. Bershadsky and H. Ooguri, *Comm. Math. Phys.* **126** (1989) 49.
- [28] F. Smirnov and L. Takhtadjan, U. Col. Boulder preprint #12 in Applied Mathematics (1990).
- [29] B. Lian and G. Zuckerman, *Phys. Lett.* **254B** (1991) 417.
- [30] V. Kazakov, *Mod. Phys. Lett.* **A4** (1989) 2125.
- [31] V. Kazakov, *Phys. Lett.* **119A** (1986) 140; D. Boulatov and V. Kazakov, *Phys. Lett.* **186B** (1987) 379.
- [32] E. Martinec, G. Moore, and N. Seiberg, *Phys. Lett.* **263B** (1991) 190.
- [33] T. Tada and M. Yamaguchi, UT-Komaba 90-17; J. DeBoer, 'Multimatrix models and the KP hierarchy', Utrecht preprint THU 91/08.
- [34] E. Brézin, G. Parisi, C. Itzykson, and J.-B. Zuber, *Comm. Math. Phys.* **59** (1979) 35; C. Itzykson and J.-B. Zuber, *J. Math Phys.* **21** (1980) 411; D. Bessis, C. Itzykson, and J.-B. Zuber, *Adv. Appl. Math.* **1** (1980) 109.
- [35] G. Harris and E. Martinec, *Phys. Lett.* **B245** (1990) 384; E. Brezin and H. Neuberger, *Nucl. Phys.* **B350** (1991) 513.
- [36] M. Staudacher, *Nucl. Phys.* **B336** (1990) 349.
- [37] P. DiFrancesco and D. Kutasov, *Nucl. Phys.* **342** (1990) 589.
- [38] P. Ginsparg, M. Goulian, M.R. Plesser, and J. Zinn-Justin, *Nucl. Phys.* **B342** (1990) 539; A. Jevicki and T. Yoneya, *Mod. Phys. Lett.* **A5** (1990) 1615; T. Yoneya, 'Toward a canonical formalism of non-perturbative 2d gravity', U. Tokyo preprint UT-Komaba 91-8.
- [39] S. Chadha, G. Mahoux, and M.L. Mehta, *J. Phys. A* **14** (1981) 579.
- [40] I. Gelfand and L. Dikii, *Funct. Anal. and Appl.* **10** (1976) 259; *Russ. Math. Surveys* **30:5** (1975) 77.

- [41] K. Apfeldorf, *Nucl. Phys.* **B360** (1991) 480.
- [42] F. David, 'What is the intrinsic geometry of 2d quantum gravity?', Rutgers preprint RU-91-25.

STRING THEORY IN TWO DIMENSIONS

Igor R. Klebanov^{*}

Joseph Henry Laboratories

Princeton University

Princeton, New Jersey 08544

ABSTRACT

I review some of the recent progress in two-dimensional string theory, which is formulated as a sum over surfaces embedded in one dimension.

1. INTRODUCTION.

These notes are an expanded version of the lectures I gave at the 1991 ICTP Spring School of Theoretical Physics. Here I have attempted to review, from my own personal viewpoint, some of the exciting developments in two-dimensional string theory that have taken place over the last year and a half. Because of the multitude of new results, and since the field is still developing rapidly, a comprehensive review must await a later date. These notes are mainly devoted to the matrix model approach¹ that has truly revolutionized the two-dimensional Euclidean quantum gravity. Recently this approach has led to the exact solution of quantum gravity coupled to conformal matter systems with $c \leq 1$.²⁻⁶

These lectures are about the $c = 1$ model^{5,6} that is both the richest physically and among the most easily soluble. This model is defined by the sum over geometries embedded in 1 dimension. The resulting string theory is, however, 2-dimensional because the dynamical conformal factor of the world sheet geometry acts as an extra hidden string coordinate.⁷ In fact, this is the maximal bosonic

^{*} Supported in part by DOE grant DE-AC02-76WRO3072 and by an NSF Presidential Young Investigator Award.

string theory that is well-defined perturbatively. If the dimensionality is increased any further, then a tachyon appears in the string spectrum and renders the theory unstable. The matrix model maps the theory of surfaces embedded in 1 dimension into a non-relativistic quantum mechanics of free fermions,⁸ from which virtually any quantity can be calculated to all orders in the genus expansion. The second-quantized field theory of non-relativistic fermions can be regarded here as the exact string field theory. A transformation of this theory to a rather simple interacting boson representation^{9,10} will also be given. I will argue that the 2-dimensional string theory is the kind of toy model which possesses a remarkably simple structure, and at the same time incorporates some of the physics of string theories embedded in higher dimensions. The simplicity of the theory is apparent in the matrix model approach, but much of it remains obscure and mysterious from the point of view of the continuum Polyakov path integral. Until we develop a better insight into the “miracles” of the matrix models, we may be sure that our understanding of string physics is very incomplete.

These notes are mainly a review of published papers, but I also included a few new observations and results. In section 2 I review the formulation of quantum gravity coupled to a scalar field as a sum over discretized random surfaces embedded in 1 dimension. I will show that the matrix quantum mechanics generates the necessary statistical sum. In section 3 I exhibit the reduction to free fermions and define the double-scaling limit.⁵ The sum over continuous surfaces is calculated to all orders in the genus expansion. In section 4 I show how to calculate the exact string amplitudes using the free fermion formalism. Section 5 is devoted to the continuum path integral formalism¹¹ where the Liouville field acts as an extra dimension of string theory. Some of the exact matrix model results will be reproduced, but this approach still falls far short of the power of the matrix model. In section 6 I discuss the special states which exist in the spectrum in addition to the massless “tachyon”. These states, occurring only at integer momenta, are left-overs of the transverse excitations of string theory. Remarkably, they generate a chiral $W_{1+\infty}$ algebra. In section 7 I use the matrix model to formulate the exact string field theory both in the fermionic and in bosonic terms. I present some manifestly finite bosonic calculations which are in perfect agreement with

the fermionic ones. In section 8 I discuss the new physical effects that arise when the random surfaces are embedded in a circle of radius R : the $R \rightarrow 1/R$ duality and its breaking due to Kosterlitz-Thouless vortices.⁶ This model can be regarded either as a compactified Euclidean string theory or as a Minkowski signature string theory at finite temperature. In section 9 I show that in the matrix quantum the effects of vortices are implemented by the states in the non-trivial representations of $SU(N)$.^{6,12} Including only the $SU(N)$ singlet sector gives the vortex-free continuum limit where the sum over surfaces respects $R \rightarrow 1/R$ duality.⁶ In section 10 I use the thermal field theory of non-relativistic fermions to find exact amplitudes of the compactified string theory.¹³ Finally, in section 11 I use the matrix chain model to solve a string theory with a discretized embedding dimension.⁶ Remarkably, when the lattice spacing is not too large, this theory is exactly equivalent to string theory with a continuous embedding.

2. DISCRETIZED RANDOM SURFACES AND MATRIX QUANTUM MECHANICS.

In this section I will introduce the discretized approach to summation over random surfaces embedded in one dimension and its matrix model implementation. If we parametrize continuous surfaces by coordinates (σ_1, σ_2) , then the Euclidean geometry is described by the metric $g_{\mu\nu}(\sigma)$, and the embedding – by the scalar field $X(\sigma)$. Thus, the theory of random surfaces in one embedding dimension is equivalent to 2-d quantum gravity coupled to a scalar field. In the Euclidean path integral approach to such a theory, we have to sum over all the compact connected geometries and their embeddings,

$$\mathcal{Z} = \sum_{\text{topologies}} \int [Dg_{\mu\nu}][DX] e^{-S}. \quad (1)$$

The integration measure is defined modulo reparametrizations, *i.e.* different functions describing the same geometry-embedding in different coordinates should not

be counted separately. We assume the simplest generally covariant massless action,

$$S = \frac{1}{4\pi} \int d^2\sigma \sqrt{g} \left(\frac{1}{\alpha'} g^{\mu\nu} \partial_\mu X \partial_\nu X + \Phi R + 4\lambda \right). \quad (2)$$

In 2 dimensions the Einstein term is well-known to give the Euler characteristic, the topological invariant which depends only on the genus of the surface h (the number of handles),

$$\frac{1}{4\pi} \int d^2\sigma \sqrt{g} R = 2(1 - h). \quad (3)$$

Thus, the weighting factor for a surface of genus h is $(\exp \Phi)^{2h-2}$. Since the sum over genus h surfaces can be thought of as a diagram of string theory with h loops, we identify the string coupling constant as $g_0 = e^\Phi$.

The main problem faced in the calculation of the Euclidean path integral of eq. (1) is the generally covariant definition of the measure for the sum over metrics $[Dg_{\mu\nu}]$ and of the cut-off. One may attempt to do this directly in the continuum. Considerable success along this route has been achieved when metrics are described in the light-cone gauge,¹⁴ or in the more traditional conformal gauge.¹⁵ We will carry out some comparisons with the continuum approach in the course of the lectures.

The main subject of this section is a different approach to summing over geometries, which has so far proven to be far more powerful than the continuum methods. In this approach one sums over discrete approximations to smooth surfaces, and then defines the continuum sum by taking the lattice spacing to zero.¹ We may, for instance, choose to approximate surfaces by collections of equilateral triangles of side a . A small section of such a triangulated surface is shown in fig. 1. The dotted lines show the lattice of coordination number 3, which is dual to the triangular lattice. Each face of the triangular lattice is thought of as flat, and the curvature is entirely concentrated at the vertices. Indeed, each vertex I of the triangular lattice has a conical singularity with deficit angle $\frac{\pi}{3}(6 - q_I)$, where q_I is the number of triangles that meet at I . Thus, the vertex I has a δ -function of curvature with positive, zero, or negative strength if $q_I < 6$, $= 6$, > 6 respectively. Such a distribution of curvature may seem like a “poor man’s version” of geometry. However,

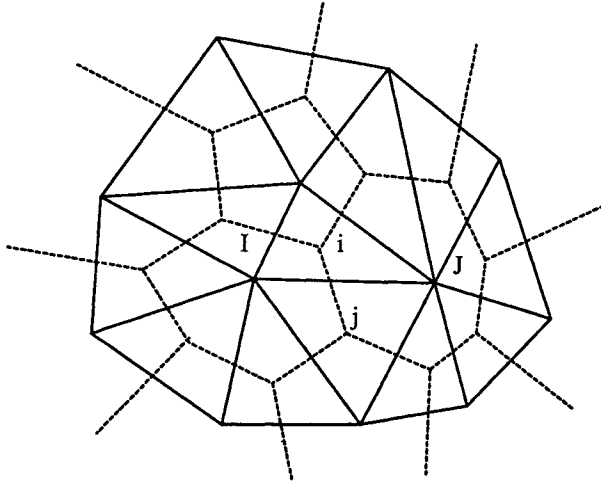


Fig. 1. A small section of triangulated surface. Solid lines denote the triangular lattice Λ , and dotted lines – the dual lattice $\tilde{\Lambda}$.

in the continuum limit the size of each face becomes infinitesimal, and we may define smoothed out curvature by averaging over many triangles. In this way, the continuum field for the metric should appear similarly to how the quantum field description emerges in the continuum limit of the more familiar statistical systems, such as the Ising model, the XY model, etc. Later on, we will show some strong indications that the discretized approach to summing over geometries is indeed equivalent to the continuum field-theoretic approach.

The essential assumption of the discretized approach is that sum over genus h geometries, $\int [Dg_{\mu\nu}]_h$, may be defined as the sum over all distinct lattices Λ , with the lattice spacing subsequently taken to zero. $\int [DX]$ is then defined as the integral over all possible embeddings of the lattice Λ in the real line. For simplicity, the lattices Λ may be taken to be triangular, but admixture of higher polygons should not affect the continuum limit. In order to complete the definition of the discretized approach, we need to specify the weight attached to each configuration. This can be defined by simply discretizing the action of eq. (2) and counting each distinct configuration with weight e^{-S} . We will find it convenient to specify the embedding coordinates X at the centers of the triangles, *i.e.* at the vertices i of the dual lattice.

Then the discretized version of $\frac{1}{4\pi\alpha'} \int d^2\sigma \sqrt{g} g^{\mu\nu} \partial_\mu X \partial_\nu X$ is $\sim \sum_{\langle ij \rangle} (X_i - X_j)^2$, where the sum runs over all the links $\langle ij \rangle$ of the dual lattice. Similarly,

$$\int d^2\sigma \sqrt{g} \rightarrow \frac{\sqrt{3}}{4} a^2 V \quad (4)$$

where V is the number of triangles, or, equivalently, the number of vertices of the dual lattice $\tilde{\Lambda}$. Thus, the discretized version of the path integral (1) is

$$\mathcal{Z}(g_0, \kappa) = \sum_h g_0^{2(h-1)} \sum_{\Lambda} \kappa^V \prod_{i=1}^V \int dX_i \prod_{\langle ij \rangle} G(X_i - X_j) \quad , \quad (5)$$

where Λ are all triangular lattices of genus h , $\kappa \sim \exp(-\sqrt{3}\lambda a^2/4\pi)$, and (for some choice of α') $G(X) = \exp(-\frac{1}{2}X^2)$. If the continuum limit of eq. (5) indeed describes quantum gravity coupled to a scalar field, then there should exist a whole universality class of link factors G , of which the gaussian is only one representative, that result in the same continuum theory.

A direct evaluation of the lattice sum (5) seems to be a daunting task, which is still outside the numerical power of modern computers. Fortunately, there exists a remarkable trick which allows us to exactly evaluate sums over surfaces of any topology using only analytic tools. As was first noted by Kazakov and Migdal,¹⁶ a statistical sum of the form (5) is generated in the Feynman graph expansion of the quantum mechanics of a $N \times N$ hermitian matrix.* Consider the Euclidean path integral

$$Z = \int D^{N^2} \Phi(x) \exp \left[-\beta \int_{-T/2}^{T/2} dx \operatorname{Tr} \left(\frac{1}{2} \left(\frac{\partial \Phi}{\partial x} \right)^2 + U(\Phi) \right) \right] \quad . \quad (6)$$

where x is the Euclidean time and $U = \frac{1}{2\alpha'} \Phi^2 - \frac{1}{3!} \Phi^3$. The parameter β enters as

* Similar tools work for other simple matter systems coupled to two-dimensional gravity. For example, as discussed in other lectures in this volume, in the case of pure gravity the sum over discretized surfaces is generated simply by an integral over an $N \times N$ hermitian matrix.

the inverse Planck constant. By a rescaling of Φ eq. (6) can be brought to the form

$$Z \sim \int D^{N^2} \Phi(x) \exp \left[-N \int_{-T/2}^{T/2} dx \operatorname{Tr} \left(\frac{1}{2} \left(\frac{\partial \Phi}{\partial x} \right)^2 + \frac{1}{2\alpha'} \Phi^2 - \frac{\kappa}{3!} \Phi^3 \right) \right] , \quad (7)$$

where $\kappa = \sqrt{N/\beta}$ is the cubic coupling constant. The connection with the statistical sum (5) follows when we develop the graph expansion of Z in powers of κ . The Feynman graphs all have coordination number 3, and are in one-to-one correspondence with the dual lattices $\tilde{\Lambda}$ of the discretized random surfaces (fig. 1). The lattices Λ dual to the Feynman graphs can be thought of as the basic triangulations. One easily obtains the sum over all connected graphs $\ln Z$,

$$\lim_{T \rightarrow \infty} \ln Z = \sum_h N^{2-2h} \sum_{\Lambda} \kappa^V \prod_{i=1}^V \int dx_i \prod_{\langle ij \rangle} e^{-|x_i - x_j|/\alpha'} . \quad (8)$$

This is precisely of the same form as eq. (5) which arises in two-dimensional quantum gravity! The Euclidean time x assumes the role of the embedding coordinate X in eq. (5). We note that the role of the link factor G is played by the one-dimensional Euclidean propagator. Only for this exponential G can we establish the exact equivalence with the matrix model. This does not pose a problem, however, as we will find plenty of evidence that the continuum limit of the model (5) with the exponential G indeed describes quantum gravity coupled to a scalar, eq. (1). It is evident from (8) that the parameter α' sets the scale of the embedding coordinate. In fact, we have normalized α' so that in the continuum limit it will precisely coincide with the definition in eq. (2). Whenever α' is not explicitly mentioned, its value has been set to 1.

Further comparing eqs. (8) and (5), we note that the size of the matrix N enters as $1/g_0$. Let us show why. Each vertex contributes a factor $\sim N$, each edge (propagator) $\sim 1/N$, and each face contributes N because there are as many index loops as there are faces. Thus, each graph is weighted by N^{V-E+F} which, by Euler's theorem, equals N^{2-2h} . The expansion of the free energy of the matrix quantum mechanics in powers of $1/N^2$ automatically classifies surfaces according

to topology. It would seem that, in order to define a theory with a finite string coupling, it is necessary to consider finite N which, as we will see, is associated with severe difficulties. Fortunately, this naive expectation is false: in the continuum limit the “bare” string coupling $1/N$ becomes infinitely multiplicatively renormalized. Thus, if N is taken to ∞ simultaneously with the world sheet continuum limit, then we may obtain a theory with a finite string coupling. This remarkable phenomenon, known as the “double-scaling limit”^{2,5}, will allow us to calculate sums over continuum surfaces of any topology.

In order to reach the continuum limit, it is necessary to increase the cubic coupling $\kappa = \sqrt{N/\beta} \sim N^0$ until it reaches the critical value κ_c where the average number of triangles in a surface begins to diverge. In this limit we may think of each triangle as being of infinitesimal area $\sqrt{3}a^2/4$, so that the whole surface has a finite area and is continuous. In the continuum limit $\kappa \rightarrow \kappa_c$ we define the cosmological constant

$$\Delta = \pi(\kappa_c^2 - \kappa^2) . \quad (9)$$

Recalling that $\kappa = \kappa_c \exp(-\sqrt{3}\lambda a^2/4\pi)$, we establish the connection between Δ and the physical cosmological constant λ : $\lambda \sim \Delta/a^2$.

Above we have sketched the connection between matrix quantum mechanics and triangulated random surfaces. It is easy to generalize this to the case where, in addition to triangles, the surfaces consist of other n -gons: we simply have to add to the matrix potential $U(\Phi)$ monomials Φ^n , $n > 3$. For consistency, the continuum limit should not be sensitive to the precise manner in which the surfaces are discretized. In the next section we will give the exact solution of the matrix model and show that the continuum limit is indeed universal.

3. MATRIX QUANTUM MECHANICS AND FREE FERMIONS.

In the previous section we established the equivalence of the sum over connected discretized surfaces to the logarithm of the path integral of the matrix quantum mechanics. On the other hand, in the hamiltonian language,

$$Z = \langle f | e^{-\beta HT} | i \rangle . \quad (10)$$

Thus, as long as the initial and final states have some overlap with the ground state, the ground state energy E_0 will dominate the $T \rightarrow \infty$ limit of eq. (10),

$$\lim_{T \rightarrow \infty} \frac{\ln Z}{T} = -\beta E_0 . \quad (11)$$

The divergence of the sum over surfaces proportional to the length of the embedding dimension arises due to the translation invariance. In order to calculate the sum over surfaces embedded in the infinite real line R^1 , all we need to find is the ground state energy of the matrix quantum mechanics. To this end we carry out canonical quantization of the $SU(N)$ symmetric hermitian matrix quantum mechanics.^{8,17} The Minkowski time lagrangian

$$L = \text{Tr} \left\{ \frac{1}{2} \dot{\Phi}^2 - U(\Phi) \right\} \quad (12)$$

is symmetric under time-independent $SU(N)$ rotations $\Phi(t) \rightarrow V^\dagger \Phi(t) V$. It is convenient to decompose Φ into N eigenvalues and $N^2 - N$ angular degrees of freedom,

$$\Phi(t) = \Omega^\dagger(t) \Lambda(t) \Omega(t) \quad (13)$$

where $\Omega \in SU(N)$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$. The only term in the lagrangian which gives rise to dependence on Ω is

$$\text{Tr} \dot{\Phi}^2 = \text{Tr} \dot{\lambda}^2 + \text{Tr} [\lambda, \dot{\Omega} \Omega^\dagger]^2 . \quad (14)$$

To identify the canonical angular degrees of freedom, the anti-hermitian traceless

matrix $\dot{\Omega}\Omega^\dagger$ can be decomposed in terms of the generators of $SU(N)$,

$$\dot{\Omega}\Omega^\dagger = \frac{i}{\sqrt{2}} \sum_{i < j} \dot{\alpha}_{ij} T_{ij} + \dot{\beta}_{ij} \tilde{T}_{ij} + \sum_{i=1}^{N-1} \dot{\alpha}_i H_i \quad (15)$$

where H_i are the diagonal generators of the Cartan subalgebra, and the other generators are defined as follows. T_{ij} is the matrix M such that $M_{ij} = M_{ji} = 1$, and all other entries are 0; $\tilde{T}_{ij}$ is the matrix M such that $M_{ij} = -M_{ji} = -i$, and all other entries are 0. Substituting eq. (15) into eqs. (14) and (12), we find

$$L = \sum_i \left(\frac{1}{2} \dot{\lambda}_i^2 + U(\lambda_i) \right) + \frac{1}{2} \sum_{i < j} (\lambda_i - \lambda_j)^2 (\dot{\alpha}_{ij}^2 + \dot{\beta}_{ij}^2) . \quad (16)$$

In deriving the hamiltonian it is crucial⁸ that the measure of integration expressed in terms of Ω and Λ assumes the form $\mathcal{D}\Phi = \mathcal{D}\Omega \prod_i d\lambda_i \Delta^2(\lambda)$, where $\Delta(\lambda)$ is the Vandermonde determinant $\prod_{i < j} (\lambda_i - \lambda_j)$. Because of the non-trivial Jacobian, the kinetic term for the eigenvalues becomes

$$-\frac{1}{2\beta^2} \sum_{i=1}^N \frac{1}{\Delta^2(\lambda)} \frac{d}{d\lambda_i} \Delta^2(\lambda) \frac{d}{d\lambda_i} . \quad (17)$$

It is left as an exercise for the reader to show that this can be rewritten as

$$-\frac{1}{2\beta^2 \Delta(\lambda)} \sum_i \frac{d^2}{d\lambda_i^2} \Delta(\lambda) . \quad (18)$$

It follows that the total hamiltonian is

$$H = -\frac{1}{2\beta^2 \Delta(\lambda)} \sum_i \frac{d^2}{d\lambda_i^2} \Delta(\lambda) + \sum_i U(\lambda_i) + \sum_{i < j} \frac{\Pi_{ij}^2 + \tilde{\Pi}_{ij}^2}{(\lambda_i - \lambda_j)^2} , \quad (19)$$

where Π_{ij} and $\tilde{\Pi}_{ij}$ are the momenta conjugate to α_{ij} and β_{ij} respectively, *i.e.*, they are the generators of *left* rotations on Ω , $\Omega \rightarrow A\Omega$. Furthermore, since L is independent of $\dot{\alpha}_i$, the wave functions must obey the constraints $\Pi_i \Psi = 0$. These constraints arise because transformations $\Omega \rightarrow A\Omega$, where A is a diagonal $SU(N)$ matrix, do not change Φ .

The constraints require that only those irreducible representations of $SU(N)$ that have a state with all weights equal to zero are allowed in the matrix quantum mechanics.¹² For instance, the fundamental representations are excluded, and the simplest non-trivial representation is the adjoint. It is not hard to classify all the irreducible representations allowed in the matrix quantum mechanics, *i.e.*, the ones that have states with zero weights. In the standard notation, the Young tableaux of such representations consist of $i + j$ columns, with the numbers of boxes in the columns (from left to right)

$$(N - m_1, N - m_2, \dots, N - m_j, n_i, \dots, n_2, n_1) \quad (20)$$

where $n_1 \leq n_2 \leq \dots \leq n_i \leq N/2$ and $m_1 \leq \dots \leq m_j \leq N/2$.

Since the angular kinetic terms in (19) are positive definite, it is clear that we have to look for the ground state among the wave functions that are annihilated by Π_{ij} and $\tilde{\Pi}_{ij}$, *i.e.* in the trivial (singlet) representation of $SU(N)$. The singlet wave functions are independent of the angles Ω , and are symmetric functions of the eigenvalues, $\chi_{sym}(\lambda)$. Because of the special form of the hamiltonian (19), the eigenvalue problem $H\chi_{sym} = E\chi_{sym}$ assumes the remarkable form⁸

$$\left(\sum_{i=1}^N h_i \right) \Psi(\lambda) = E\Psi(\lambda) , \quad (21)$$

$$h_i = -\frac{1}{2\beta^2} \frac{d^2}{d\lambda_i^2} + U(\lambda_i) ,$$

where $\Psi(\lambda) = \Delta(\lambda)\chi_{sym}(\lambda)$ is by construction antisymmetric. Since the operator acting on Ψ is simply the sum of single-particle non-relativistic hamiltonians, the problem has been reduced to the physics of N free non-relativistic fermions moving in the potential $U(\lambda)$.⁸ This equivalence to free fermions is at the heart of the exact solubility of string theory in two dimensions.

The N -fermion ground state is obtained by filling the lowest N energy levels of h ,

$$E_0 = \sum_{i=1}^N \epsilon_i . \quad (22)$$

The potential arising from triangulated random surfaces is $U(\lambda) = \frac{1}{2}\lambda^2 - \frac{1}{3!}\lambda^3$, which

is unbounded from below and does not seem support any stable states. Recall, however, that in order to find the genus expansion of the sum over surfaces, we are only interested in the expansion about the classical limit $\beta \rightarrow \infty$ ($\hbar \rightarrow 0$) in powers of $1/\beta^2$. Classically, there is a continuum of bounded orbits to the left of the potential barrier. Semiclassically, these orbits are associated with energy levels (fig. 1) whose spacing is $\mathcal{O}(1/\beta)$, and whose decay time is exponentially long as $\beta \rightarrow \infty$. This instability cannot be seen in the context of the semiclassical expansion in powers of $1/\beta^2$. Our goal, therefore, is to find the expansion of the ground state energy (22), where ϵ_i are the semiclassical energy levels.

To leading order, ϵ_n are determined by the Bohr-Sommerfeld quantization condition,

$$\oint p_n(\lambda) d\lambda = \frac{2\pi}{\beta} n \quad (23)$$

where $p_n(\lambda) = \sqrt{2(\epsilon_n - U(\lambda))}$, and the integral is over a closed classical orbit. In particular, the position of the Fermi level $\mu_F = \epsilon_N$ is determined by

$$\int_{\lambda_-}^{\lambda_+} \sqrt{2(\mu_F - U(\lambda))} d\lambda = \pi \frac{N}{\beta}, \quad (24)$$

where λ_- and λ_+ are the classical turning points. Eq. (24) shows that μ_F is an increasing function of N/β . Clearly, both μ_F and E_0 become singular at the point where $N/\beta \rightarrow \kappa_c^2$ such that μ_F equals the height of the barrier μ_c . Recall that the cubic coupling of the matrix quantum mechanics is $\kappa = \sqrt{N/\beta}$, and that we expect to find a singularity in the sum over graphs when κ becomes large enough that the average number of vertices in a Feynman graph (or equivalently, of triangles in a random surface) begins to diverge. In the vicinity of this singularity the continuum limit of quantum gravity can be defined. Now we have identified the physics of this singularity in the equivalent free fermion system: it is associated with spilling of fermions over the barrier.

We are interested in the singularity of E_0 as a function of the cosmological constant $\Delta = \pi(\kappa_c^2 - \frac{N}{\beta})$. The parts of $E(\Delta)$ analytic in Δ are not universal and can be dropped. Indeed, if we calculate the sum over surfaces as a function

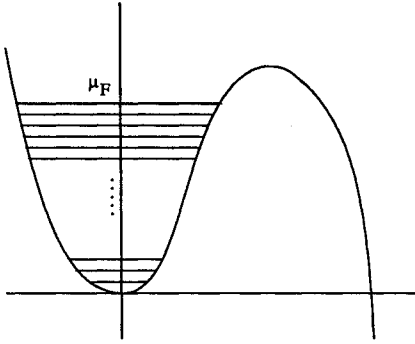


Fig. 2a) N fermions in the asymmetric potential arising directly from the triangulated surfaces.

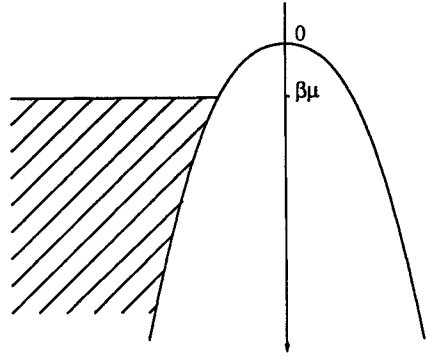


Fig. 2b) The double-scaling limit magnifies the quadratic local maximum.

of the fixed area by inversely Laplace transforming $E(\Delta)$, the analytic terms give contributions only for zero area.

Let us introduce the density of eigenvalues

$$\rho(\epsilon) = \frac{1}{\beta} \sum_n \delta(\epsilon - \epsilon_n), \quad (25)$$

in terms of which

$$\begin{aligned} \kappa^2 &= \frac{N}{\beta} = \int_0^{\mu_F} \rho(\epsilon) d\epsilon ; \\ \lim_{T \rightarrow \infty} \frac{-\ln Z}{T} &= \beta E = \beta^2 \int_0^{\mu_F} \rho(\epsilon) \epsilon d\epsilon . \end{aligned} \quad (26)$$

Differentiating the first equation, we find

$$\frac{\partial \Delta}{\partial \mu} = \pi \rho(\mu_F) , \quad (27)$$

which can be integrated and inverted to obtain $\mu = \mu_c - \mu_F$ as a function of Δ .

Differentiating the second equation, we find $\frac{\partial E}{\partial \Delta} = \frac{1}{\pi} \beta(\mu - \mu_c)$. After addition to E

of an irrelevant analytic term, $E \rightarrow E + \frac{1}{\pi}\beta\mu_c\Delta$, we find

$$\frac{\partial E}{\partial \Delta} = \frac{1}{\pi}\beta\mu \quad (28)$$

These equations show that the continuum limit of the sum over surfaces is determined by the singularity of the single particle density of states near the top of the barrier. Using the WKB approximation, it was found in ref. 16 that

$$\rho(\mu_F) = \frac{1}{\pi} \int_{\lambda_-}^{\lambda_+} \frac{d\lambda}{\sqrt{2(\mu_F - U(\lambda))}} \sim -\frac{1}{\pi} \ln \mu + \mathcal{O}(1/\beta^2) \quad (29)$$

It follows that $\Delta = -\mu \ln \mu + \mathcal{O}(1/\beta^2)$, and

$$-\beta E = \frac{1}{2\pi}(\beta\mu)^2 \ln \mu + \dots \approx \frac{1}{2\pi} \frac{\beta^2 \Delta^2}{\ln \Delta} + \dots \quad (30)$$

The critical behavior of ρ comes from the part of the integral near the quadratic maximum of the potential. In the continuum limit $\mu \rightarrow 0$, the classical trajectory at μ_F spends logarithmically diverging amount of time near the maximum. Therefore, the WKB wave functions are peaked there. In fact, we will show that, at any order in the WKB expansion, the leading singularity of ρ is entirely determined by the quadratic nature of the maximum. To see this, let us rescale the coordinates $\sqrt{\beta}(\lambda - \lambda_c) = y$, and $e = \mu_c - \epsilon$. The single particle eigenvalue problem becomes⁵

$$\left(\frac{1}{2} \frac{d^2}{dy^2} + \frac{1}{2} y^2 + \mathcal{O}(1/\sqrt{\beta}) y^3 \right) \psi(y) = \beta e \psi(y) \quad (31)$$

For a finite βe , all the terms beyond the quadratic are irrelevant in the limit $\beta \rightarrow \infty$ because they are suppressed by powers of β . This demonstrates the universality of the continuum limit: any potential possessing a quadratic maximum will yield the same sum over continuous surfaces. In effect, our rescaling has swept all the non-universal details of the potential out to infinity. This implies, for instance, that, instead of working with the potential of fig. 2, we can use the double well potential $U(\lambda) = -\frac{1}{2}\lambda^2 + \lambda^4$ (fig. 3). After rescaling $\sqrt{\beta}\lambda = y$, the Schrödinger

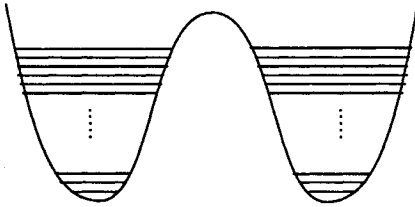


Fig. 3a) N fermions in a double well potential.

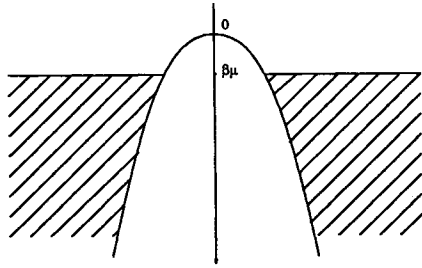


Fig. 3b) The system relevant to the double-scaling limit.

equation again reduces to the motion in the inverted harmonic oscillator. The only difference is that now fermions fill both sides of the maximum. To all orders of the WKB expansion this simply multiplies the density of states, and therefore the free energy, by a factor of 2. The double well potential does not suffer from the non-perturbative instabilities: the ground state energy can in principle be calculated even for finite N and β . We will often adopt the symmetric case because it is more convenient for calculations, remembering that we have to divide by 2 to find the sum over triangulated surfaces.

From eq. (30) we know the sum over all surfaces of spherical topology embedded in R^1 . In order to find the sums over surfaces of any topology, we need to develop a systematic WKB expansion of the ground state energy. According to eqs. (27) and (28), all we need to know is the complete expansion of the single particle density of states $\rho(\mu)$. If we consider μ such that $\beta\mu$ is $\mathcal{O}(1)$, then this problem has a remarkably simple solution,⁵ because eq. (31) shows that in this energy range the system is purely quadratic. Let us write the density of states as

$$\rho(\mu) = \text{Tr} \delta(h_0 + \beta\mu) = \frac{1}{\pi} \text{Im} \text{Tr} \left[\frac{1}{h_0 + \beta\mu - i\epsilon} \right], \quad (32)$$

$$h_0 = -\frac{1}{2} \frac{d^2}{dy^2} - \frac{1}{2} y^2.$$

We will evaluate the trace in position space in order to take advantage of the fact

that the resolvent of the hamiltonian of a simple harmonic oscillator is well known,

$$\begin{aligned} \langle y_f | \frac{1}{-\frac{1}{2}\frac{d^2}{dy^2} + \frac{1}{2}\omega^2 y^2 + \beta\mu - i\epsilon} | y_i \rangle &= \int_0^\infty dT e^{-\beta\mu T} \int_{y_i}^{y_f} \mathcal{D}y(t) e^{-\int_0^T dt \frac{1}{2}(\dot{y}^2 + \omega^2 y^2)} \\ &= \int_0^\infty dT e^{-\beta\mu T} \sqrt{\frac{\omega}{2\pi \sinh \omega T}} e^{-\omega[(y_f^2 + y_i^2) \cosh \omega T - 2y_f y_i]/2 \sinh \omega T} . \end{aligned} \quad (33)$$

In our case, the frequency ω is actually imaginary (we have an *inverted* harmonic oscillator), so we must define the resolvent by analytic continuation. We rotate $\omega \rightarrow -i$ while simultaneously rotating $T \rightarrow iT$, and the relevant resolvent is

$$\langle y_f | \frac{1}{h_0 + \beta\mu - i\epsilon} | y_i \rangle = i \int_0^\infty dT e^{-i\beta\mu T} \sqrt{\frac{-i}{2\pi \sinh T}} e^{i[(y_f^2 + y_i^2) \cosh T - 2y_f y_i]/2 \sinh T} . \quad (34)$$

Now it is easy to derive the derivative of the density of states⁵

$$\frac{\partial \rho}{\partial(\beta\mu)} = \frac{1}{\pi} \frac{\partial}{\partial(\beta\mu)} \text{Im} \int_{-\infty}^\infty dy \langle y | \frac{1}{h_0 + \beta\mu - i\epsilon} | y \rangle = \frac{1}{\pi} \text{Im} \int_0^\infty dT e^{-i\beta\mu T} \frac{T/2}{\sinh(T/2)} . \quad (35)$$

Note that only for the purposes of deriving the large $\beta\mu$ expansion can we reduce the problem to pure inverted harmonic oscillator. The integral representation (35) is not designed to correctly give the non-perturbative terms $\mathcal{O}(e^{-\beta\mu})$. In fact, such terms depend on the details of the potential away from the quadratic maximum, and are not universal. While each term in the large $\beta\mu$ expansion is related to the sum over all geometries of a certain genus, we do not yet fully understand the geometrical or physical meaning of the non-perturbative terms.

After expanding eq. (35) and integrating once, we find the complete series of corrections to the leading order estimate in eq. (29),

$$\rho(\mu) = \frac{1}{\pi} \left\{ -\ln \mu + \sum_{m=1}^\infty (2^{2m-1} - 1) \frac{|B_{2m}|}{m} (2\beta\mu)^{-2m} \right\} . \quad (36)$$

In solving eq. (27) for μ in terms of Δ , it is useful to introduce parameter μ_0

defined by

$$\Delta = -\mu_0 \ln \mu_0 = \mu_0 |\ln \mu_0|. \quad (37)$$

Then,

$$\mu = \mu_0 \left\{ 1 - \frac{1}{\ln \mu_0} \sum_{m=1}^{\infty} (2^{2m-1} - 1) \frac{|B_{2m}|}{m(2m-1)} (2\beta\mu_0)^{-2m} + \mathcal{O}\left(\frac{1}{\ln^2 \mu_0}\right) \right\}. \quad (38)$$

Integrating (28) we finally arrive at the complete genus expansion of the sum over surfaces

$$-\beta E = \frac{1}{8\pi} \left\{ (2\beta\mu_0)^2 \ln \mu_0 - \frac{1}{3} \ln \mu_0 + \sum_{m=1}^{\infty} \frac{(2^{2m+1} - 1)|B_{2m+2}|}{m(m+1)(2m+1)} (2\beta\mu_0)^{-2m} \right\}. \quad (39)$$

Clearly, the sensible way to approach the continuum limit is through “double scaling” $\beta \rightarrow \infty$ and $\mu_0 \rightarrow 0$ in such a way that $\beta\mu_0 = 1/g_{st}$ is kept fixed.^{5,6} Thus, quite remarkably, we have found closed form expressions for sums over continuous surfaces of any topology embedded in one dimension. The sum over genus h surfaces, given by the coefficient of g_{st}^{2h-2} , grows as $(2h)!$. This behavior is characteristic of closed string theories,¹⁸ while in field theory we typically find that the sum over h -loop diagrams exhibits a slower growth $\sim h!$. The badly divergent perturbation series (39) indicates that the theory is not well-defined non-perturbatively.

Eq. (39) shows explicitly how the bare string coupling $1/\beta$ gets multiplicatively renormalized. Compared to the $c < 1$ matter coupled to gravity,²⁻⁴ the new feature here is that the renormalization is not simply by a power of the cosmological constant Δ . It appears that Δ itself is multiplicatively renormalized: $\Delta \rightarrow \mu_0 \approx \Delta/|\ln \Delta|$. Another peculiarity of the $c = 1$ matter is that the sums over spherical and toroidal surfaces diverge in the continuum limit as $|\ln \mu_0|$. In section 5 we will argue that all these features have a natural explanation in the context of Liouville theory. Before we do that, however, we further exhibit the power of the matrix model by calculating some correlation functions to all orders in the topological expansion.

4. CORRELATION FUNCTIONS.

The simplest kind of matrix model operator is $\text{Tr } \Phi^n(x)$ where n is any finite integer. This operator inserts a vertex of order n in all possible places on the dual graph $\tilde{\Lambda}$. In terms of the basic lattice Λ , this amounts to cutting an n -gon hole in the surface, pinning it at the embedding coordinate x , and integrating the position of the hole all over the surface. In the continuum limit, for any finite n the size of the hole becomes infinitesimal, and it is commonly referred to as “a puncture”. The continuum expression for the operator above should be roughly

$$\int d^2\sigma \sqrt{g} \delta(X(\sigma) - x) . \quad (40)$$

By Fourier transforming, we derive the continuum translation of a related operator

$$\int dx e^{iqx} \text{Tr } \Phi^n(x) \rightarrow \int d^2\sigma \sqrt{g} e^{iqX} , \quad (41)$$

which resembles the basic “tachyon” operator of string theory. Below we outline the procedure for calculating the correlation functions of such operators in the matrix model. In section 5 we show that the results can indeed be reproduced in the string theoretic formalism.

Since the matrix model has been reduced to a system of free fermions, it is convenient to perform calculations in the formalism of non-relativistic second quantized field theory. The second quantized fermion field is defined as

$$\hat{\psi}(\lambda, x) = \int_{-\infty}^{\infty} d\nu e^{i\nu x} a_{\epsilon}(\nu) \psi^{\epsilon}(\nu, \lambda) , \quad (42)$$

where $\psi^{\epsilon}(\nu, \lambda)$ are the single fermion wave functions of energy ν , and the associated annihilation operators $a_{\epsilon}(\nu)$ satisfy $\{a_{\epsilon}^{\dagger}(\nu), a_{\epsilon}(\nu')\} = \delta_{\epsilon, \epsilon'} \delta(\nu - \nu')$. Setting $\lambda = \sqrt{2}y$ in eq. (31), we find that $\psi^{\epsilon}(\nu, \lambda)$ are solutions of the Whittaker equation

$$\left(\frac{d^2}{d\lambda^2} + \frac{\lambda^2}{4} \right) \psi = \nu \psi . \quad (43)$$

and can therefore be expressed in terms of Whittaker functions $W(\nu, \lambda)$.¹⁹ Their spectrum is continuous and doubly degenerate. Here ϵ denotes the parity of the wavefunction $\psi^{\epsilon}(\nu, \lambda)$ and repeated ϵ are summed over.

The Euclidean second-quantized field theory with chemical potential $\beta\mu$ is defined by the action

$$S = \int_{-\infty}^{\infty} d\lambda \int_{-\infty}^{\infty} dx \, \hat{\psi}^\dagger \left(-\frac{d}{dx} + \frac{d^2}{d\lambda^2} + \frac{\lambda^2}{4} + \beta\mu \right) \hat{\psi} \quad (44)$$

where x is the Euclidean time. The ground state of the system (fig. 3b) satisfies

$$\begin{aligned} a_\epsilon(\nu)|\beta\mu\rangle &= 0, & \nu < \beta\mu, \\ a_\epsilon^\dagger(\nu)|\beta\mu\rangle &= 0, & \nu > \beta\mu. \end{aligned} \quad (45)$$

Consider correlation functions of operators of the type

$$\mathcal{O}(q) = \int dx e^{iqx} \text{Tr} f(\Phi(x)), \quad (46)$$

where f is any function. In the second quantized formalism, they translate into

$$\mathcal{O}(q) = \int dx e^{iqx} \int d\lambda f(\lambda) \hat{\psi}^\dagger \hat{\psi}(x, \lambda). \quad (47)$$

Thus, to find the connected correlation functions of any set of such operators, we need to calculate

$$G(q_1, \lambda_1; \dots; q_n, \lambda_n) = \prod_{i=1}^n \int dx_i e^{iq_i x_i} \langle \beta\mu | \hat{\psi}^\dagger \hat{\psi}(x_1, \lambda_1) \dots \hat{\psi}^\dagger \hat{\psi}(x_n, \lambda_n) | \beta\mu \rangle_c. \quad (48)$$

Application of Wick's theorem reduces this to the sum over one-loop diagrams with fermion bilinear insertions in all possible orders around the loop. A convenient formula for the Euclidean Green function is¹⁹

$$\begin{aligned} S^E(x_1, \lambda_1; x_2, \lambda_2) &= \int_{-\infty}^{\infty} d\nu e^{-(\nu - \beta\mu)\Delta x} \{ \theta(\Delta x) \theta(\nu - \beta\mu) - \theta(-\Delta x) \theta(\beta\mu - \nu) \} \times \\ \psi^\epsilon(\nu, \lambda_1) \psi^\epsilon(\nu, \lambda_2) &= \int_{-\infty}^{\infty} \frac{dp}{2\pi} e^{-ip\Delta x} \int_{-\infty}^{\infty} d\nu \frac{i}{p + i(\nu - \beta\mu)} \psi^\epsilon(\nu, \lambda_1) \psi^\epsilon(\nu, \lambda_2) \\ &= i \int_{-\infty}^{\infty} \frac{dp}{2\pi} e^{-ip\Delta x} \int_0^{sgn(p)\infty} ds e^{-sp + i\beta\mu s} \langle \lambda_1 | e^{2isH} | \lambda_2 \rangle \end{aligned} \quad (49)$$

where $\Delta x = x_1 - x_2$, and we have used¹⁹

$$\int_{-\infty}^{\infty} d\nu e^{-i\nu s} \psi^\epsilon(\nu, \lambda_1) \psi^\epsilon(\nu, \lambda_2) = \langle \lambda_1 | e^{2isH} | \lambda_2 \rangle = \frac{1}{\sqrt{-4\pi i \operatorname{sh} s}} \exp \left(-\frac{i}{4} \left[\frac{\lambda_1^2 + \lambda_2^2}{\operatorname{th} s} - \frac{2\lambda_1 \lambda_2}{\operatorname{sh} s} \right] \right) \quad (50)$$

with $H = 1/2p^2 - 1/8\lambda^2$. With the Green function (49), after integration over the loop momentum, Moore derived the representation

$$\frac{\partial}{\partial \mu} G(q_1, \lambda_1; \dots; q_n, \lambda_n) = i^{n+1} \delta(\sum q_i) \sum_{\sigma \in \Sigma_n} \int_{-\infty}^{\infty} d\xi e^{i\mu\xi} \int_0^{\epsilon_1 \infty} ds_1 \dots \int_0^{\epsilon_{n-1} \infty} ds_{n-1} e^{-s_1 Q_1^\sigma - \dots - s_{n-1} Q_{n-1}^\sigma} \langle \lambda_{\sigma(1)} | e^{2is_1 H} | \lambda_{\sigma(2)} \rangle \dots \langle \lambda_{\sigma(n)} | e^{2i(\xi - \sum_{i=1}^{n-1} s_i) H} | \lambda_{\sigma(1)} \rangle \quad (51)$$

where $Q_k^\sigma = q_{\sigma(1)} + \dots + q_{\sigma(k)}$, and $\epsilon_k = \operatorname{sgn}[Q_k^\sigma]$. In principle, any correlation function of operators of type (46) can be found by integrating G . A convenient way of calculating correlations of puncture operators is to first introduce the operator that creates a finite boundary of length l , with momentum q injected into the boundary¹⁹

$$O(l, q) = \int dx e^{iqx} \operatorname{Tr} e^{-l\Phi(x)}. \quad (52)$$

The puncture operator should be the leading term in the small l expansion of $O(l, q)$. The correlations of puncture operators $P(q)$ can be extracted according to

$$\left\langle \prod_i O(l_i, q_i) \right\rangle \sim \prod_i l_i^{|q_i|} \left\langle \prod_j P(q_j) \right\rangle. \quad (53)$$

The details of the calculations are highly technical and can be found in ref. 19. Here we will simply state the results for the two, three, and four point functions.

$$\langle P(q) P(k) \rangle = -\delta(q+k) [\Gamma(1-|q|)]^2 \mu^{|q|} \left[\frac{1}{|q|} - \frac{1}{24(\beta\mu)^2} (|q|-1)(q^2-|q|-1) + \dots \right]. \quad (54)$$

$$\begin{aligned}
\langle P(q_1)P(q_2)P(q_3) \rangle &= \delta \left(\sum q_i \right) \prod_{i=1}^3 \left(\Gamma(1 - |q_i|) \mu^{\frac{|q_i|}{2}} \right) \frac{1}{\beta \mu} \left[1 \right. \\
&- \frac{1}{24(\beta \mu)^2} (|q_3| - 1)(|q_3| - 2)(q_1^2 + q_2^2 - |q_3| - 1) \\
&+ \prod_{r=1}^4 (|q_3| - r) \left(3(q_1^4 + q_2^4) + 10q_1^2 q_2^2 - 10(q_1^2 + q_2^2)|q_3| - 5(q_1 - q_2)^2 + 12|q_3| + 7 \right) \\
&\left. \times \frac{1}{5760(\beta \mu)^2} + \dots \right],
\end{aligned} \tag{55}$$

where $q_1, q_2 > 0$.

$$\begin{aligned}
\langle P(q_1)P(q_2)P(q_3)P(q_4) \rangle &= -\delta \left(\sum q_i \right) \prod_{i=1}^4 \left(\Gamma(1 - |q_i|) \mu^{\frac{|q_i|}{2}} \right) \times \\
&\frac{1}{2(\beta \mu)^2} \left[(|q_1 + q_2| + |q_1 + q_3| + |q_1 + q_4| - 2) + \mathcal{O}(1/(\beta \mu)^2) \right].
\end{aligned} \tag{56}$$

The new feature of the $c = 1$ correlators, compared to theories with $c < 1$, is that they have genuine divergences occurring at quantized values of the external momenta.²⁰ In this sense, the $c = 1$ model is the most similar to critical string theory where the amplitudes are well-known to have divergences associated with the production of on-shell physical particles. The remarkable feature of the $c = 1$ amplitudes is that all the divergences factorize into external leg factors.

Because of the divergences, eqs. (54)-(56) are strictly valid only if none of q_i are integer. If any of the momenta have integer values, then additional terms need to be taken into account, which leads to cancellation of the infinity.²⁰ The physical meaning of this is simple: in the matrix model there is an explicit ultraviolet and infrared cut-off, and therefore there are no genuine zeroes or poles of the correlation functions. Instead, they are regularized according to $1/0 \rightarrow |\ln \mu_0|$, and $0 \rightarrow 1/|\ln \mu_0|$. Note that in eqs. (54)-(56) μ can be replaced by μ_0 . Thus, the correlation functions scale as powers of μ_0 , and not as powers of Δ found in all the $c < 1$ theories coupled to gravity.²⁻⁴ This provides further evidence that, for $c = 1$, a renormalization of the cosmological constant takes place. In the next section we sketch a Liouville theory explanation of some of the new effects peculiar to $c = 1$.

5. LIOUVILLE GRAVITY COUPLED TO $c = 1$ MATTER.

In this section we outline the continuum approach to evaluating the sum over surfaces (1). In order to integrate over all metrics, in 2 dimensions we may pick the conformal gauge $g_{\mu\nu} = \hat{g}_{\mu\nu}(\tau)e^{-\phi}$ where $\hat{g}_{\mu\nu}$ depends on a finite number of modular parameters collectively denoted as τ . For spherical topology, there are no moduli, and we may choose $\hat{g}_{\mu\nu} = \delta_{\mu\nu}$. Eq. (2) shows that classically the Liouville field ϕ is non-dynamical. It is well-known, however,¹¹ that the kinetic term for ϕ is induced through quantum effects arising from dependence on ϕ of the path integral measure. Additional modification of the dynamics of ϕ has been proposed by David and by Distler and Kawai,¹⁵ who argued that we can simplify the integration measure at the expense of further renormalization of the action for ϕ such that the ghost, X and ϕ fields combine into a conformally invariant theory with net conformal anomaly equal to zero. Following their approach, we assume that the original path integral (1) is transformed in the conformal gauge into

$$\begin{aligned} Z &= \sum_h g_0^{2h-2} \int [d\tau][DbDc][D\phi][DX] e^{-S_{b,c}-S} , \\ S &= \frac{1}{4\pi} \int d^2\sigma \sqrt{\hat{g}} \left(\partial_\mu X \partial^\mu X + \partial_\mu \phi \partial^\mu \phi - 4\hat{R}\phi + 4\Delta e^{-2\phi} \right) , \end{aligned} \quad (57)$$

and $S_{b,c}$ is the standard free action for the b and c ghosts. The exponential interaction term originates from the cosmological term in the action (2).^{*}

With the action (57), the simplest set of conformally invariant operators are

$$T(q) = \frac{1}{\pi\beta} \int d^2\sigma \sqrt{\hat{g}} e^{iqX + \epsilon(q)\phi} \quad (58)$$

These operators inject embedding momentum q into the world sheet. They should be thought of as the Liouville theory implementation of the operators (41). In a

^{*} Polchinski has argued²¹ that, for $c = 1$, the interaction term is not simply an exponential. We will instead work with the exponential interaction, and will show that it is also possible to explain correlation functions within this framework. Our argument does not substantially differ from Polchinski's, but we prefer working with the exponential because it was shown by Curtwright and Thorn²² that with this form of interaction the conformal invariance can be maintained exactly in the continuum limit.

theory with $\Delta = 0$, $T(q)$ is conformally invariant for $\epsilon(q) = -2 \pm |q|$. For $\Delta > 0$, however, it was shown in ref. 23, 24 that only the operator with $\epsilon(q) = -2 + |q|$ exists. This conclusion agrees with the scaling of correlation functions in the matrix model. Let us now use the Liouville formalism to calculate the spherical correlation functions.

As discovered in ref. 25, in performing the path integral it is convenient to decompose $\phi = \phi_0 + \tilde{\phi}$, and integrate first over the zero mode ϕ_0 . Then, since the interaction is purely exponential, we obtain

$$\left\langle \prod_{i=1}^N T(q_i) \right\rangle_h = g_0^{2h-2} \delta\left(\sum q_i\right)^{\frac{1}{2}} \left(\frac{\Delta}{\pi}\right)^s \Gamma(-s) \int [d\tau][D\tilde{X}][D\tilde{\phi}][DbDc] e^{-S_b, c - S_0} \left(\int d^2\sigma \sqrt{\hat{g}} e^{-2\phi} \right)^s \prod_i \frac{1}{\pi\beta} \int d^2\sigma \sqrt{\hat{g}} e^{iq_i X + (-2 + |q_i|)\phi} \quad (59)$$

where

$$S_0 = \frac{1}{4\pi} \int d^2\sigma \sqrt{\hat{g}} \left(\partial_\mu \tilde{X} \partial^\mu \tilde{X} + \partial_\mu \tilde{\phi} \partial^\mu \tilde{\phi} \right) \quad (60)$$

and $s = 2(1 - h) - N + \frac{1}{2} \sum_{i=1}^N |q_i|$. The factor $\Gamma(-s)$ indicates that correlation functions with non-negative integer s have a divergence from the volume of ϕ_0 . This explains the divergences in the 0-point functions on the sphere and torus found in eq. (39). Now, the difficulty is to deal with the insertion of $T(0)^s$. Goulian and Lee²⁶ proposed to calculate eq. (59) first for integer s , where the rules are perfectly clear, and then to define it for non-integer s by analytic continuation. This gives answers in agreement with the matrix model.^{26,27}

Now we show how the multiplicative renormalization $\Delta \rightarrow \mu_0$ takes place. Although it seems that eq. (59) scales as Δ^s , the free field correlator multiplying it is, formally, of order 0^s . For integer s this can be shown through direct calculation. Consider, for instance, the tree level correlator of M tachyons with $q_M < 0$, $q_1, \dots, q_{M-1} > 0$. If we impose the condition $s = 0$, then $q_M = 2 - M$. The integral for the coefficient of $\frac{1}{2}\Gamma(0)$ can be performed explicitly²⁷ and gives

$$\frac{\beta^{2-M}}{(M-3)!} \prod_{i=1}^{M-1} \frac{\Gamma(1 - |q_i|)}{\Gamma(|q_i|)}, \quad (61)$$

where we have set $g_0^2 = \pi^3 \beta^2$. If we now send $q_1, q_2, \dots, q_m \rightarrow 0$, while keeping $\sum_1^{M-1} q_i = M - 2$, we obtain a tachyon correlator with $s = m$. From eq. (61) we obtain

$$\left\langle \prod_{i=m+1}^M T(q_i) \right\rangle = \frac{1}{2} \beta^{2-M+m} \Gamma(-m) \Delta^m 0^m \frac{1}{(M-3)!} \prod_{i=m+1}^{M-1} \frac{\Gamma(1-|q_i|)}{\Gamma(|q_i|)} \quad (62)$$

The physical meaning of the zero of order m is that the “tachyon” decouples at zero momentum. This is a sign of the fact that the “tachyon” is actually a massless particle in two-dimensional string theory. We will show this more explicitly later on.

Now we would like to argue that in a cut-off theory 0^s should be replaced by $(1/|\ln \mu_0|)^s$. If the cut-off is imposed, then the Liouville zero mode is enclosed in a box of size $\sim |\ln \mu_0|$. Thus, the lowest accessible value of the “Liouville energy” $2 + \epsilon(q)$ is $\sim 1/|\ln \mu_0|$, and there is no complete decoupling of $T(0)$. Hence, the amplitudes should scale as $(\Delta/|\ln \mu_0|)^s = \mu_0^s$, which is the sought for renormalization of the cosmological constant. We have identified the physical origin of this renormalization: the decoupling of the massless particle at zero momentum. •

The continuation of eq. (62) to non-integer s gives the N -point tachyon amplitudes with $q_N < 0$, $q_1, \dots, q_{N-1} > 0$,²⁷

$$\left\langle \prod_{i=1}^N T(q_i) \right\rangle = \frac{1}{2} \beta^{2-N} \prod_{i=1}^N \frac{\Gamma(1-|q_i|)}{\Gamma(|q_i|)} \left(\frac{d}{d\mu_0} \right)^{N-3} \mu_0^{s+N-3}. \quad (63)$$

This agrees with the matrix model results (54)-(56) if the matrix model and Liouville theory operators are related by momentum dependent normalization factor

$$T(q) = \frac{1}{\Gamma(|q|)} P(q) \quad (64)$$

With this correspondence, we can use the matrix model results to obtain tachyon correlation functions to all orders in the genus expansion. This is remarkable because direct Liouville theory calculations of correlation functions have not yet been performed even for genus 1. For $h > 3$ we do not know the region of modular

integration and could not perform a direct Liouville calculation even in principle. From eq. (54) we find, for instance,

$$\langle T(q)T(k) \rangle = -\delta(q+k) \left[\frac{\Gamma(1-|q|)}{\Gamma(|q|)} \right]^2 \mu_0^{|q|} \left[\frac{1}{|q|} - \frac{1}{24(\beta\mu)^2} (|q|-1)(q^2-|q|-1) + \dots \right]. \quad (65)$$

A natural continuation of this formula to $q = 0$ is to replace $\lim_{q \rightarrow 0} |q|$ by $1/|\ln \mu_0|$. Then we obtain,

$$\langle T(0)T(0) \rangle \sim \frac{\partial^2}{\partial \Delta^2} E. \quad (66)$$

This is a simple consistency check: differentiation of the path integral with respect to Δ inserts a zero-momentum tachyon.

To conclude this section, we will describe the only $h > 0$ calculation that has been performed to date. This is the sum over toroidal surfaces with no insertions.²⁸ The reason this calculation is relatively straightforward is that $s = 0$, and the complicated insertion of $T(0)^s$ is absent. Sending $s \rightarrow 0$ in eq. (59), and remembering the renormalization $\Delta \rightarrow \mu_0$, we obtain the free field path integral

$$\mathcal{Z}_1 = -\frac{1}{2} \ln \mu_0 \int_{\mathcal{F}} d^2 \tau \int [D\tilde{X}][D\tilde{\phi}][DbDc] e^{-S_0} \quad (67)$$

where $\tau = \tau_1 + i\tau_2$ is the modular parameter of the torus, and $\mathcal{F}$ denotes the fundamental region of the modular group: $\tau_2 > 0$, $|\tau| > 1$, $-\frac{1}{2} \leq \tau_1 < \frac{1}{2}$. This path integral is standard, and we find

$$\mathcal{Z}_1(R/\sqrt{\alpha'}) = -\frac{1}{2} \ln \mu_0 \int_{\mathcal{F}} d^2 \tau \left(\frac{|\eta(q)|^4}{2\tau_2} \right) (2\pi\sqrt{\tau_2}|\eta(q)|^2)^{-1} Z(R/\sqrt{\alpha'}, \tau, \bar{\tau}), \quad (68)$$

where η is the Dedekind function and $q = e^{2\pi i \tau}$. In the above integrand, the first term is from the ghost determinant, the second is from the Liouville determinant, and $Z(R/\sqrt{\alpha'}, \tau, \bar{\tau})$ is the partition function of the scalar field compactified on a

circle,

$$Z(R/\sqrt{\alpha'}, \tau, \bar{\tau}) = \frac{R}{\sqrt{\alpha'}} \frac{1}{\sqrt{\tau_2} |\eta(q)|^2} \sum_{n,m=-\infty}^{\infty} \exp\left(-\frac{\pi R^2 |n - m\tau|^2}{\alpha' \tau_2}\right). \quad (69)$$

The double sum is over the soliton winding numbers about the two non-trivial cycles of the torus. From eqs. (68) and (69), we obtain

$$\begin{aligned} \mathcal{Z}_1(R/\sqrt{\alpha'}) &= -\frac{1}{2} \ln \mu_0 \frac{R}{4\pi\sqrt{\alpha'}} F(R/\sqrt{\alpha'}), \\ F(R/\sqrt{\alpha'}) &= \int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} \sum_{n,m} \exp\left(-\frac{\pi R^2 |n - m\tau|^2}{\alpha' \tau_2}\right). \end{aligned} \quad (70)$$

The η -functions cancel out, and we end up only with the contributions of the zero modes. This is due to the absence of particles corresponding to the transverse string excitations, and reflects the two-dimensional nature of this string theory.^{*} Remarkably, the integral in eq. (70) can be easily performed using the trick of ref. 30. The idea is to trade the m -sum over the winding sectors of the string for a sum over many inequivalent fundamental regions which together cover the strip $-\frac{1}{2} \leq \tau_1 < \frac{1}{2}$ in the upper half-plane. In 26 dimensions this gives a representation of the string free energy in terms of the free energies of all the modes of the string³¹. Similarly, eq. (70) becomes

$$F(R/\sqrt{\alpha'}) = \int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} + 2 \int_0^\infty \frac{d\tau_2}{\tau_2^2} \sum_{k=1}^{\infty} \exp\left(-\frac{\pi R^2 k^2}{\alpha' \tau_2}\right). \quad (71)$$

The second term is simply the temperature-dependent one-loop free energy for a single massless boson in 2 dimensions, expressed in the proper time representation. The first term is the one-loop cosmological constant of the massless boson, which has been automatically supplied with a ‘stringy’ ultraviolet cut-off: here the τ

^{*} The isolated transverse states at discrete momenta, to be discussed in section 6, do not affect the torus partition function.²⁹

integral is over the fundamental region, not over the strip. Performing the integrals in eq. (71), we find $F(R/\sqrt{\alpha'}) = \frac{\pi}{3}(1 + \alpha'/R^2)$ and

$$\mathcal{Z}_1(R/\sqrt{\alpha'}) = -\frac{1}{24}\left(\frac{R}{\sqrt{\alpha'}} + \frac{\sqrt{\alpha'}}{R}\right)\ln \mu_0. \quad (72)$$

As we will show in section 9, this answer agrees with the matrix model result. This calculation gives us hope that the Liouville path integral exactly describes sums over geometries beyond the tree level.

6. THE SPECIAL OPERATORS.

The most striking feature of the tachyon correlation functions is the occurrence of the external leg factors $\frac{\Gamma(1-|q_i|)}{\Gamma(|q_i|)}$, which contain factorized poles whenever any of the momenta approaches an integer value. Although this exact factorization is not yet fully understood, it was noted in ref. 20 that the poles are related to the presence of other physical states in the theory besides the tachyons (58). These states are remnants of the transverse excitations of the string and occur only at integer q .^{20,32,33}

The origin of these states can be traced to the pure $c = 1$ conformal field theory where there are special primary fields of the form³⁴

$$V_{J,m}(\partial X, \partial^2 X, \dots) e^{2miX(z)} \quad (73)$$

with conformal weight J^2 . These states form $SU(2)$ multiplets with spin J and magnetic number m . The connection with $SU(2)$ becomes apparent once we note that the states (73) are the full set of primary fields in the compact $c = 1$ theory with the self-dual radius $R = 1$, where there is a well-known level 1 $SU(2)$ current algebra. For each spin J , $V_{J,\pm J} = 1$, which gives a tachyon operator. Other members of the multiplet can be constructed by applying raising and lowering operators $\frac{1}{2\pi i} \oint dz \exp(\pm 2iX(z))$. After coupling to gravity, we can “dress” the

special $c = 1$ operators to obtain new $(1, 0)$ operators of the form

$$\mathcal{V}_{J,m}(z) = V_{J,m} e^{2miX} e^{2(J-1)\phi}(z) . \quad (74)$$

Recently, it was established that these operators are related to a $W_{1+\infty}$ algebra.^{35†} Undoubtedly, this algebra plays an important role in determining the properties of the theory.

In the matrix model, the physical $(1, 1)$ operator

$$\int d^2 z \mathcal{V}_{J,m}(z) \bar{\mathcal{V}}_{J,m}(\bar{z}) \quad (75)$$

was, up to normalization, identified with^{10,36}

$$\int_{-\infty}^{\infty} dx e^{2miX} (\lambda - \lambda_c)^{2J} \quad (76)$$

where λ_c is the coordinate of the quadratic maximum. Using this identification, the correlation functions of all operators with $m = 0$ were calculated in ref. 10. Recently, it was shown^{37,38} that the operators (76) generate a $W_{1+\infty}$ algebra. The connection of this algebra with the algebra of special operators in Liouville theory has been established by E. Witten.³⁵

The simplest new operator is the zero-momentum dilaton

$$\mathcal{V}_{1,0} \bar{\mathcal{V}}_{1,0} = \partial X \bar{\partial} X . \quad (77)$$

Thus, of the full dilaton field of the critical string theory only its zero-momentum part remains here. Similarly, the higher J excitations correspond to remnants of the higher mass particles. The connection of the special operators to the divergences in the correlation functions is provided by the fusion rules. There are terms $\sim 1/|z - w|^2$ in the fusion rules of the physical operators (75). Integration over one operator near another can thus produce logarithmic divergences.

† A related observation was made by I. Klebanov and A. Polyakov.

Let us elaborate on this mechanism using, as an example, the tachyon 2-point function.²⁰ The poles in it occur when $|q|$ is integer, *i.e.* when tachyon operators become members of the $SU(2)$ multiplets of the special operators. The operator product expansion of two integer momentum tachyons with $n > 0$ is

$$e^{inX+(-2+|n|)\phi}(z, \bar{z})e^{-inX+(-2+|n|)\phi}(w, \bar{w}) \sim \dots + \frac{1}{|z-w|^2} \mathcal{V}_{|n|-1,0} \bar{\mathcal{V}}_{|n|-1,0} + \dots \quad (78)$$

Integrating over z near w , we obtain a logarithmic divergence for any integer $|q| > 0$ due to the appearance of a physical operator in the fusion rule. This explains the infinite sequence of poles in the two-point function.

7. STRING FIELD THEORY FROM THE MATRIX MODEL.

The success of the Liouville approach shows that the theory of surfaces embedded in one dimension can also be viewed as critical string theory embedded in *two* dimensions.^{7,21} The field ϕ , which starts out as the conformal factor of the world sheet metric, assumes the role of another embedding dimension similar to X . Indeed, the path integral (57) can be thought of as the sigma model for the bosonic string propagating in two-dimensional Euclidean target space with the metric $G_{\mu\nu} = \delta_{\mu\nu}$, the dilaton background $D(\phi) = -4\phi$, and the tachyon background $T(\phi) = \Delta e^{-2\phi}$. The two-dimensionality of the string theory has many important physical consequences.

First of all, the mass of the particle corresponding to motion of the string in its ground state, “the tachyon” is

$$m_T^2 = \frac{2-D}{6} \quad (79)$$

Thus, $D = c + 1 = 2$ is the “critical dimension of non-critical string theory” where “the tachyon” is exactly massless. Indeed, in the calculation of the Liouville path integral on the torus, we found a massless two-dimensional boson propagating around the loop. From the decoupling at zero momentum, observed in the study of the tachyon correlation functions, we deduce that the theory is invariant under constant shifts of the tachyon field.

The second expected feature of two-dimensionality is the absence of transverse oscillation modes of the string. Indeed, if we could pass to the light-cone gauge, then all transverse oscillations would be eliminated and the entire spectrum of the string theory would consist of one massless field. In reality, we cannot pick the light-cone gauge because of the lack of translation invariance in the ϕ -coordinate. As a result, the spectrum of the string contains the transverse excitations constructed in section 6. These states are not full-fledged fields, however, because they only occur at discrete momenta $|q| = n$. Indeed, the string one-loop calculation of section 5 showed that there are no massive transverse *fields* in the theory. Thus, at the minimal level, we should be able to formulate the string field theory with a single massless field, perhaps coupled to an infinite number of quantum mechanical degrees of freedom.

In this section we show that a string field theory with the features anticipated above can be derived directly from the matrix quantum mechanics. This further strengthens the connection of $c = 1$ quantum gravity with string theory in $D = 2$. The matrix quantum mechanics directly reduces to the exact string field theory, which is the second-quantized field theory of non-relativistic fermions discussed in section 4. In this section we will simply recast this formalism so that some of its physical features become more transparent. First, we will show that the non-relativistic field theory can be expressed in the double-scaling limit as a theory of free quasi-relativistic chiral fermions.^{10,39} These chiral fermions have the kinetic term that is relativistic to order g_{st}^0 , but receives translationally non-invariant corrections of order g_{st} . Further, we will carry out the conventional bosonization⁴⁰ of this fermionic hamiltonian, and obtain an interacting theory of massless bosons in $D = 2$ first derived by Das and Jevicki with somewhat different methods.⁹ In conclusion, we will perform a few manifestly finite calculations in the bosonic formalism to show that it is fully equivalent to the original representation in terms of non-relativistic fermions.

7.1. CHIRAL FERMIONS.

In order to exhibit our general method, we first consider fermions moving in an arbitrary potential $U(\lambda)$. When we later take the double-scaling limit, as expected it will depend only on the existence of a quadratic local maximum of $U(\lambda)$. The second quantized hamiltonian for a system of free fermions with Planck constant $1/\beta$ is

$$\hat{h} = \int d\lambda \left\{ \frac{1}{2\beta^2} \frac{\partial \Psi^\dagger}{\partial \lambda} \frac{\partial \Psi}{\partial \lambda} + U(\lambda) \Psi^\dagger \Psi - \mu_F (\Psi^\dagger \Psi - N) \right\}, \quad (80)$$

where μ_F is the Lagrange multiplier necessary to fix the total number of fermions to equal N . As usual, it will be adjusted so as to equal the Fermi level of the N fermion system, $\mu_F = \epsilon_N$. The fermion field has the expansion

$$\Psi(\lambda, t) = \sum_i \alpha_i \psi_i(\lambda) e^{-i\beta \epsilon_i t}, \quad (81)$$

where $\psi_i(\lambda)$ are the single particle wave functions and α_i are the respective annihilation operators. The time t is now taken to be Minkowskian.

All the known exact results of the $c = 1$ matrix model have been derived from this nonrelativistic field theory. An important feature of this theory is that it is two-dimensional: the field $\Psi(\lambda, t)$ depends on t , and also on the eigenvalue coordinate λ . This is the simplest way to see how the hidden dimension, originating from the Liouville mode of quantum gravity, emerges in the matrix model.

Note that, as $\beta \rightarrow \infty$, the single-particle spectrum is approximately linear near the Fermi surface,

$$\epsilon_n \approx \mu_F + \frac{1}{\beta} n\omega + \mathcal{O}\left(\frac{1}{\beta^2}\right). \quad (82)$$

This suggests that the behavior of excitations near the Fermi surface is quasi-relativistic. The spectrum of relativistic fermions in a box is, of course, exactly linear. We will therefore find non-relativistic corrections arising from the violations of the linearity of ϵ_n due to higher orders of the semiclassical expansion. As the Fermi level approaches the maximum, $\mu = \mu_c - \mu_F \rightarrow 0$, the spacing of the spectrum vanishes, $\omega \approx \pi/|\ln \mu|$. Thus, we expect that the size of the box in which the quasi-relativistic fermions are enclosed diverges $\sim |\ln \mu|$.

Let us introduce new fermionic variables Ψ_L and Ψ_R ,

$$\Psi(\lambda, t) = \frac{e^{-i\mu_F t}}{\sqrt{2v(\lambda)}} \left[e^{-i\beta \int^\lambda d\lambda' v(\lambda') + i\pi/4} \Psi_L(\lambda, t) + e^{i\beta \int^\lambda d\lambda' v(\lambda') - i\pi/4} \Psi_R(\lambda, t) \right] , \quad (83)$$

where $v(\lambda)$ is the velocity of the classical trajectory of a particle at the Fermi level in the potential $U(\lambda)$,

$$v(\lambda) = \frac{d\lambda}{d\tau} = \sqrt{2(\mu_F - U(\lambda))} . \quad (84)$$

We substitute (83) into (80) to derive the hamiltonian in terms of Ψ_L and Ψ_R . We drop all terms which contain rapidly oscillating exponentials of the form $\exp [\pm 2i\beta \int^\lambda v(x)dx]$, since these give exponentially small terms as $\beta \sim N \rightarrow \infty$ and do not contribute to any order of semiclassical perturbation theory. For the same reason we can restrict the coordinate λ to lie between the turning points of the classical motion. For $\mu > 0$, we will consider the case where the fermions are localized on one side of the barrier. If we considered the symmetric case of fig. 3, we would have found two identical worlds decoupled from each other to all orders of the semiclassical expansion. We will also discuss $\mu < 0$ where the Fermi level is above the barrier, and the fermions can be found on both sides.

After some algebra we find

$$\begin{aligned} \mathcal{H} = 2\beta\hbar = & \int_0^{T/2} d\tau \left[i\Psi_R^\dagger \partial_\tau \Psi_R - i\Psi_L^\dagger \partial_\tau \Psi_L + \frac{1}{2\beta v^2} (\partial_\tau \Psi_L^\dagger \partial_\tau \Psi_L + \partial_\tau \Psi_R^\dagger \partial_\tau \Psi_R) \right. \\ & \left. + \frac{1}{4\beta} (\Psi_L^\dagger \Psi_L + \Psi_R^\dagger \Psi_R) \left(\frac{v''}{v^3} - \frac{5(v')^2}{2v^4} \right) \right] , \end{aligned} \quad (85)$$

where $v' \equiv dv/d\tau$. Here we see that the natural spatial coordinate, in terms of which the fermion has a standard Dirac action to leading order in β , is τ – the classical time of motion at the Fermi level – rather than λ . An important feature of the hamiltonian is that it does not mix Ψ_L and Ψ_R ,

$$\mathcal{H} = \mathcal{H}_L + \mathcal{H}_R . \quad (86)$$

The only mixing between the different chiralities is through the boundary conditions. In order to determine the boundary conditions, consider the leading semi-

classical expression for $\Psi(\lambda, 0)$,

$$\begin{aligned}\Psi(\lambda, 0) &= \sum_{n>0} \psi_n a_n + \sum_{m \leq 0} \psi_m b_m^\dagger, \\ \psi_n &= \frac{2}{\sqrt{T} v_n} \cos\left(\beta \int^\lambda d\lambda' v_n(\lambda') - \frac{\pi}{4}\right) \\ v_n &= \sqrt{2(\epsilon_n - U(\lambda))},\end{aligned}\tag{87}$$

where we have relabeled $\alpha_i = a_i$, $i > 0$, $\alpha_i = b_i^\dagger$, $i \leq 0$, and i is the number of the fermion energy level starting from the Fermi level μ_F . Expanding ϵ_n about μ_F , we find

$$\begin{aligned}\Psi(\lambda, 0) &= \frac{1}{\sqrt{2v}} \left[e^{i\beta \int^\lambda v(\lambda') d\lambda' - i\pi/4} \Psi_R + e^{-i\beta \int^\lambda v(\lambda') d\lambda' + i\pi/4} \Psi_L \right], \\ \Psi_R &= \sqrt{\frac{2}{T}} \left(\sum_{n>0} e^{2\pi i n \tau / T} a_n + \sum_{m \leq 0} e^{2\pi i m \tau / T} b_m^\dagger \right), \\ \Psi_L &= \sqrt{\frac{2}{T}} \left(\sum_{n>0} e^{-2\pi i n \tau / T} a_n + \sum_{m \leq 0} e^{-2\pi i m \tau / T} b_m^\dagger \right).\end{aligned}\tag{88}$$

Thus, Ψ_R and Ψ_L are expressed in terms of a single set of fermionic oscillators, a_n and b_m . Semiclassically, they satisfy the boundary conditions

$$\Psi_R(\tau = 0) = \Psi_L(\tau = 0), \quad \Psi_R(\tau = \frac{T}{2}) = \Psi_L(\tau = \frac{T}{2}).\tag{89}$$

These ensure that the fermion number current not flow out of the finite interval, *i.e.*, that $\bar{\Psi}(\tau) \gamma_1 \Psi(\tau) = \Psi_R^\dagger \Psi_R - \Psi_L^\dagger \Psi_L$ vanish at the boundary. They also guarantee that Ψ_R and Ψ_L are not independent fields and that we are including the correct number of degrees of freedom. The issue of dynamics at the boundary is quite subtle, however, since the corrections to the relativistic hamiltonian in eq. (85) blow up precisely at the boundary. The problem is that the semiclassical approximation breaks down at the points where $v(\tau) = 0$ even as $\beta \rightarrow 0$. One possibility of dealing with this breakdown is to carefully regularize the physics at the boundary points. This has lead to some success.^{41,42} We will suggest another method, which is problem-free from the beginning: to approach double-scaling limit with $\mu < 0$. As we will show, an equivalent procedure is to work with $\mu > 0$ but to construct the theory in terms of $\Psi(p, t)$, where p is the momentum conjugate to λ .

We have succeeded in mapping the collection of N nonrelativistic fermions, which describe the eigenvalues of Φ , onto an action which, to leading order, is just the two-dimensional Dirac action with rather standard bag-like boundary conditions. However, the $\frac{1}{\beta}$ corrections in eq. (85) cannot be disregarded in the double scaling limit because near the quadratic maximum $v(\tau) = \sqrt{2\mu}\sinh(\tau)$. In the double scaling limit the surviving correction to the relativistic hamiltonian is of order $g_{st} = 1/(\beta\mu)$,

$$\mathcal{H}_{g_{st}} = \frac{1}{4\beta\mu} \int_0^\infty \frac{d\tau}{\sinh^2(\tau)} \left[|\partial_\tau \Psi_i|^2 + \frac{1}{2} |\Psi_i|^2 \left(1 - \frac{5}{2} \coth^2(\tau)\right) \right], \quad (90)$$

where i runs over L and R . This correction does not change the non-interacting nature of the fermions, but it does render the fermion propagator non-standard.

In the double-scaling limit the boundary at $T/2$ recedes to infinity and is irrelevant. In fact, the correct expression for the hamiltonian would follow had we replaced $U(\lambda)$ by $-\frac{1}{2}\lambda^2$. However, the hamiltonian (90) still suffers from the divergence at $\tau = 0$. In view of this, let us modify the problem to approach the double-scaling limit with $\mu < 0$. Since string perturbation theory is in powers of $(\beta\mu)^{-2}$, analytic continuation to positive μ should not be problematic. Now $v^2(\tau) = 2|\mu| \operatorname{ch}^2 \tau$, and we find

$$\begin{aligned} \mathcal{H} = & \int_{-\infty}^{\infty} d\tau \left[i\Psi_R^\dagger \partial_\tau \Psi_R - i\Psi_L^\dagger \partial_\tau \Psi_L + \frac{1}{4\beta|\mu| \operatorname{ch}^2 \tau} (\partial_\tau \Psi_L^\dagger \partial_\tau \Psi_L + \partial_\tau \Psi_R^\dagger \partial_\tau \Psi_R) \right. \\ & \left. + \frac{1}{8\beta|\mu| \operatorname{ch}^2 \tau} \left(1 - \frac{5}{2} \operatorname{th}^2 \tau\right) (\Psi_L^\dagger \Psi_L + \Psi_R^\dagger \Psi_R) \right]. \end{aligned} \quad (91)$$

By choosing $\mu < 0$ we have eliminated two problems at once. First, now the hamiltonian has no divergence at $\tau = 0$. Second, there is now no relevant boundary at all, Ψ_L and Ψ_R do not mix, and there is no need to impose boundary conditions. The fact that the two chiralities do not mix to all orders of the string perturbation theory is remarkable. Its impact on the scattering processes will be discussed after we introduce the bosonized formalism.

7.2. THE BOSONIC FORMALISM

In this section we will explicitly confirm the expectation that the space-time picture of the $c = 1$ string theory involves the dynamics of one massless scalar field in two dimensions. To this end we will bosonize the fields Ψ_L and Ψ_R following the standard bosonization rules for Dirac fermions.⁴⁰ A two dimensional free massless Dirac fermion is equivalent to a single free massless scalar boson. In our case, although the fermions are free, they are not truly relativistic beyond the semiclassical limit. This will give rise to interaction terms in the equivalent bosonic field theory.

To bosonize the system we replace the fermion fields by⁴⁰

$$\begin{aligned}\Psi_L &= \frac{1}{\sqrt{2\pi}} : \exp \left[i\sqrt{\pi} \int_{\tau}^{\tau} (P - X') d\tau' \right] : , \\ \Psi_R &= \frac{1}{\sqrt{2\pi}} : \exp \left[i\sqrt{\pi} \int_{\tau}^{\tau} (P + X') d\tau' \right] : ,\end{aligned}\tag{92}$$

where X is a massless two-dimensional periodic scalar field, and P is its canonically conjugate momentum. The normal ordering in eq. (92) is in terms of the conventionally defined creation and annihilation operators, which we will utilize for the explicit calculations of section 6.3. To convert eq. (85), we make use of the following easily derived expressions

$$\begin{aligned}: \Psi_L^\dagger \partial_\tau \Psi_L - \Psi_R^\dagger \partial_\tau \Psi_R : &= \frac{i}{2} : P^2 + (X')^2 : \\ : \Psi_L^\dagger \Psi_L + \Psi_R^\dagger \Psi_R : &= -\frac{X'}{\sqrt{\pi}} \\ : \partial_\tau \Psi_L^\dagger \partial_\tau \Psi_L + \partial_\tau \Psi_R^\dagger \partial_\tau \Psi_R : &= -\sqrt{\pi} : P X' P + \frac{1}{3} (X')^3 + \frac{1}{6\pi} X''' : .\end{aligned}\tag{93}$$

Substituting these into eq. (85), we find

$$\begin{aligned}: \mathcal{H} : &= \frac{1}{2} \int_0^{T/2} d\tau : \left[P^2 + (X')^2 - \frac{\sqrt{\pi}}{\beta v^2} (P X' P + \frac{1}{3} (X')^3 + \frac{1}{6\pi} X''') - \right. \\ &\quad \left. \frac{1}{2\beta\sqrt{\pi}} X' \left(\frac{v''}{v^3} - \frac{5(v')^2}{2v^4} \right) \right] : .\end{aligned}\tag{94}$$

If we integrate by parts and discard the boundary terms, this reduces to [10]

$$: \mathcal{H} := \frac{1}{2} \int_0^{T/2} d\tau : \left[P^2 + (X')^2 - \frac{\sqrt{\pi}}{\beta v^2} (P X' P + \frac{1}{3} (X')^3) - \frac{1}{2\beta\sqrt{\pi}} X' \left(\frac{v''}{3v^3} - \frac{(v')^2}{2v^4} \right) \right] : . \quad (95)$$

The boundary conditions obeyed by the field X are determined by those of Ψ , eq. (89), which ensure that fermion number not flow out of the τ box. As we have shown, the current density $\bar{\Psi}(\tau)\gamma_1\Psi(\tau) = \Psi_R^\dagger\Psi_R - \Psi_L^\dagger\Psi_L$ vanishes at the boundary. Since this density is proportional to $\partial_t X(t, \tau)$, we deduce that $X(t, 0)$ and $X(t, \frac{T}{2})$ must be constant, i.e., X satisfies Dirichlet boundary conditions. The constraint on the total fermion number requires that $X(t, 0) - X(t, \frac{T}{2}) = 0$. We are free to choose $X(t, 0) = X(t, \frac{T}{2}) = 0$. These boundary conditions eliminate all the winding and momentum modes of X . As a result, there is no need to worry about the periodic nature of X : it acts like an ordinary scalar field in a box with Dirichlet boundary conditions.

The hamiltonian (95) and the boundary condition agree with the collective field approach.^{9,41} Physically, the massless field X describes small fluctuations of the Fermi surface. In the double-scaling limit the hamiltonian reduces to (for $\mu > 0$)

$$: \mathcal{H} := \frac{1}{2} \int_0^\infty d\tau : \left[P^2 + (X')^2 - \frac{\sqrt{\pi}}{2\beta\mu \operatorname{sh}^2 \tau} (P X' P + \frac{1}{3} (X')^3) - \frac{1 - \frac{3}{2} \operatorname{cth}^2 \tau}{12\beta\mu\sqrt{\pi} \operatorname{sh}^2 \tau} X' \right] : . \quad (96)$$

This hamiltonian, as its fermionic counterpart, suffers from a divergence at $\tau = 0$. This divergence was regularized in ref. 41 with zeta-function techniques. Alternatively, we may take the double-scaling limit with $\mu < 0$. Then

$$: \mathcal{H} := \frac{1}{2} \int_0^\infty d\tau : \left[P^2 + (X')^2 - \frac{\sqrt{\pi}}{2\beta|\mu| \operatorname{ch}^2 \tau} (P X' P + \frac{1}{3} (X')^3) - \frac{1 - \frac{3}{2} \operatorname{th}^2 \tau}{12\beta|\mu|\sqrt{\pi} \operatorname{ch}^2 \tau} X' \right] : , \quad (97)$$

and the sickness at $\tau = 0$ has disappeared. The bosonized theory of non-relativistic fermions has a remarkable structure. In the double-scaling limit there is a single cubic interaction term of order g_{st} , and a tadpole of order g_{st} . Therefore, we

can develop an expansion of correlation functions in powers of g_{st} using conventional perturbation theory. As we will show, this expansion reproduces the genus expansion of the string amplitudes.

7.3. SCATTERING AMPLITUDES.

Scattering amplitudes of the X -quanta can be related to the Euclidean correlation functions in the matrix model.⁴³ We will exhibit this relation for $\mu > 0$ ³⁸ (the $\mu < 0$ case works similarly). The finite boundary operator $O(l, q)$ from eq. (52) can be translated into

$$\frac{1}{2} \int dx e^{iqx} \int_0^\infty d\tau e^{-l\lambda(\tau)} : \Psi_L^\dagger \Psi_L + \Psi_R^\dagger \Psi_R(\tau, t) : , \quad (98)$$

where $\lambda(\tau)$ is the classical trajectory at the Fermi level. Upon bosonization, we find

$$O(l, q) \sim \int dx e^{iqx} \int d\tau e^{-l\lambda(\tau)} \partial_\tau X \sim i \int_{-\infty}^\infty dk F(k, l) k \tilde{X}(q, k) \quad (99)$$

where

$$\begin{aligned} F(k, l) &= \int_0^\infty d\tau e^{-l\lambda(\tau)} \cos(k\tau) , \\ X(x, \tau) &= \int dx e^{-iqx} \int dk \sin(k\tau) X(q, k) . \end{aligned} \quad (100)$$

Evaluating $F(k, l)$ with $\lambda(\tau) = \sqrt{2\mu} \operatorname{ch} \tau$, we find³⁸

$$F(k, l) = K_{ik}(l\sqrt{2\mu}) = \frac{\pi}{2 \sin(ik\pi)} (I_{-ik}(l\sqrt{2\mu}) - I_{ik}(l\sqrt{2\mu})) . \quad (101)$$

For small z , we will replace

$$I_\nu(z) \rightarrow (z/2)^\nu \frac{1}{\Gamma(\nu+1)} . \quad (102)$$

In calculating the Euclidean correlation functions, each operator will be connected to the rest of the Feynman graph by the propagator $1/(q^2 + k^2)$. We will deform

the k -integral of eq. (99) for each external leg, and pick up the residue of the propagator pole. There is a subtlety here: each K -function has to be split into a sum of two I -functions as in eq. (101), and the allowed sense of deformation for the two is opposite. For I_{-ik} we pick up the pole at $k = i|q|$, while for I_{ik} – the pole at $k = -i|q|$. As a result, we obtain the amputated on-shell Euclidean amplitude times a factor for each external leg, which for small l is

$$(l\sqrt{\mu/2})^{|q|} \frac{\pi}{\sin(\pi|q|)\Gamma(1+|q|)} = -(l\sqrt{\mu/2})^{|q|} \Gamma(-|q|) . \quad (103)$$

Therefore, the correlation function of puncture operators is

$$\left\langle \prod_{i=1}^N P(q_i) \right\rangle \sim \prod_{j=1}^N (-\mu^{|q_j|/2} \Gamma(-|q_j|)) A(q_1, \dots, q_N) \quad (104)$$

where A is the Euclidean continuation of a scattering amplitude of N X -quanta. The same factor for each external leg appears in the direct fermionic calculations, (54)-(56). The lesson from this rather technical exercise is that the Das-Jevicki field theory assembles the correlation functions in a rather remarkable way. Each scattering amplitude of the X -quanta has to be multiplied by the external leg factors which arise because of the unusual form (99) of the external operators.⁴³ On the other hand, these factors containing the poles have an important string theoretic meaning and certainly cannot be discarded. Thus, we are still missing the precise interpretation of the Das-Jevicki field X in the string theory with Euclidean signature.

After continuation of string amplitudes to the Minkowski signature, $X \rightarrow it$, the connection with the Das-Jevicki field theory is more straightforward. The Euclidean external leg factor for the operator $T(q)$ is $-\mu^{|q|/2} \Gamma(-|q|)/\Gamma(|q|)$. Upon continuation $|q| \rightarrow -iE$, this factor turns into a pure phase¹³

$$-\exp\left(-i\frac{E}{2}\ln\mu\right) \frac{\Gamma(iE)}{\Gamma(-iE)} \quad (105)$$

which does not affect any scattering cross-sections. Since the external leg phases are unobservable, they can be absorbed in the definition of the vertex operators.

Thus, we find that the Minkowskian scattering amplitudes of tachyons are given by the corresponding amplitudes of the X -quanta.

In fact, we may consider two kinds of scattering experiments. The usual scattering involves colliding right-moving and left-moving wave packets. In this theory such an experiment yields trivial results since $\mathcal{H} = \mathcal{H}_L + \mathcal{H}_R$ ⁴³, and the wave packets pass through each other with no influence or time delay.⁴⁴ Thus, there is no conventional “bulk” scattering, whose rate is finite per unit spatial volume when plane waves are being scattered.

We can, however, consider another kind of scattering.⁴⁴ For $\mu > 0$ we may prepare a left-moving wave packet incident on the boundary and wait for it to reflect. This can be interpreted as the scattering of n left-moving particles into m right-moving particles. (Recall that the number of massless particles in one spatial dimension is a subtle concept that needs to be carefully defined.) For $\mu < 0$, a left- (right-) moving wave packet stays left- (right-) moving to all orders of perturbation theory in g_{st} . However, now there are two asymptotic regions, $\tau \rightarrow \pm\infty$, and the wave packet undergoes some deformation as it passes through the interaction region near $\tau = 0$. This deformation can be interpreted as a change of state (and number) of particles. It is this “non-bulk” scattering,^{44,41,42} whose rate is not proportional to the spatial volume, that gives, upon Euclidean continuation, the matrix model correlators. To demonstrate this explicitly, we will now calculate tree-level amplitudes for $2 \rightarrow 1$ and $2 \rightarrow 2$ particle scattering.

We will work with the $\mu < 0$ hamiltonian of eq. (97) and show that the calculations are manifestly finite. Since the chiralities do not mix, we will consider scattering of right-moving particles described by the hamiltonian

$$\mathcal{H}_R = \frac{1}{4} \int_{-\infty}^{\infty} : \left[(P - X')^2 + \frac{\sqrt{\pi}}{6\beta|\mu| \operatorname{ch}^2 \tau} (P - X')^3 + \frac{1 - \frac{3}{2} \operatorname{th}^2 \tau}{12\beta|\mu| \sqrt{\pi} \operatorname{ch}^2 \tau} (P - X') \right] : . \quad (106)$$

Following ref. 41, we will perform our calculation in the hamiltonian formalism. This is advantageous here because bosonization maps the fermionic hamiltonian into the normal-ordered bosonic one. Since fermionic calculations are finite, we expect that the normal ordering will remove all the divergences in the bosonic

approach. In other words, the hamiltonian approach provides a definite set of rules on how to handle the bosonized theory so that it is perfectly equivalent to the original fermionic theory.

We introduce the canonical oscillator basis

$$\begin{aligned} X(t, \tau) &= \int_{-\infty}^{\infty} \frac{dk}{\sqrt{4\pi|k|}} \left(a(k) e^{i(k\tau - |k|t)} + a^\dagger(k) e^{-i(k\tau - |k|t)} \right), \\ P(t, \tau) &= -i \int_{-\infty}^{\infty} \frac{dk}{\sqrt{4\pi|k|}} |k| \left(a(k) e^{i(k\tau - |k|t)} - a^\dagger(k) e^{-i(k\tau - |k|t)} \right), \end{aligned} \quad (107)$$

such that $[a(k), a^\dagger(k')] = \delta(k - k')$. The hamiltonian assumes the form $\mathcal{H}_R = H_2 + H_3 + H_1$.

$$\begin{aligned} H_2 &= \int_0^\infty dk k a^\dagger(k) a(k), \\ H_3 &= \frac{i}{24\pi\beta|\mu|} \int_0^\infty dk_1 dk_2 dk_3 \sqrt{k_1 k_2 k_3} [f(k_1 + k_2 + k_3) a(k_1) a(k_2) a(k_3) - \\ &\quad 3f(k_1 + k_2 - k_3) : a(k_1) a(k_2) a^\dagger(k_3) :] + h.c., \\ H_1 &= -\frac{i}{48\pi\beta|\mu|} \int_0^\infty dk \sqrt{k} g(k) (a(k) - a^\dagger(k)), \end{aligned} \quad (108)$$

where

$$\begin{aligned} f(k) &= \int_{-\infty}^{\infty} d\tau \frac{1}{\text{ch}^2 \tau} e^{ik\tau} = \frac{\pi k}{\text{sh}(\pi k/2)}, \\ g(k) &= \int_{-\infty}^{\infty} d\tau \frac{1 - \frac{3}{2} \text{th}^2 \tau}{\text{ch}^2 \tau} e^{ik\tau} = \frac{\pi(k^3 + 2k)}{4 \text{sh}(\pi k/2)}. \end{aligned} \quad (109)$$

Now we calculate the S -matrix,

$$S = 1 - 2\pi i \delta(E_i - E_f) T. \quad (110)$$

Each right-moving massless particle has energy equal to momentum, $E = k$. Consider the amplitude for two particles with momenta k_1 and k_2 to scatter into a

single particle of momentum k_3 ,

$$S(k_1, k_2; k_3) = -2\pi i \delta(E_1 + E_2 - E_3) \sqrt{E_1 E_2 E_3} < k_3 | H_{int} | k_1 k_2 > \quad (111)$$

where $|k_1 k_2 > = a^\dagger(k_1) a^\dagger(k_2) |0 >$. Substituting the cubic interaction term, we find

$$S(E_1, E_2; E_3) = -g_{st} \delta(E_1 + E_2 - E_3) E_1 E_2 E_3 . \quad (112)$$

The fact that this amplitude describes “non-bulk” scattering is evident from the absence of a separate delta-function for momentum conservation. In order to relate this to the matrix model three-puncture correlation function, we perform Euclidean continuation $E_j \rightarrow i|q_j|$. Including the external leg factors according to eq. (104), the result is in agreement with the tree-level contribution to eq. (55).

Now we proceed to the more complicated case of non-forward scattering of particles of momenta k_1 and k_2 into particles of momenta k_3 and k_4 . The amplitude is given by second-order perturbation theory,

$$T(k_1, k_2; k_3, k_4) = \sqrt{E_1 E_2 E_3 E_4} \sum_i \frac{< k_3 k_4 | H_{int} | i > < i | H_{int} | k_1 k_2 >}{E_1 + E_2 - E_i} , \quad (113)$$

where the sum runs over all the intermediate states i . The s -channel contribution is easily seen to be⁴¹

$$S^{(s)}(k_1, k_2; k_3, k_4) = -i \frac{g_{st}^2}{8\pi} \prod_{j=1}^4 E_j \int_0^\infty dk \left(\frac{k f^2(k_1 + k_2 - k)}{k_1 + k_2 - k + i\epsilon} - \frac{k f^2(k_1 + k_2 + k)}{k_1 + k_2 + k - i\epsilon} \right) , \quad (114)$$

where the first term arises from the one-particle intermediate state, and the second one – from the five-particle intermediate state. Eq. (114) can be expressed as an integral over k from $-\infty$ to ∞ . The t - and u -channel contributions assume a similar form. Summing over the three channels, we obtain⁴¹

$$S(k_1, k_2; k_3, k_4) = -i \frac{\pi g_{st}^2}{8} \delta(E_1 + E_2 - E_3 - E_4) \prod_{j=1}^4 E_j (F(p_s) + F(p_t) + F(p_u)) , \quad (115)$$

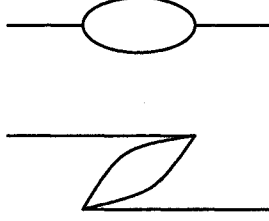


Fig. 4a) The one-loop $\mathcal{O}(g_{st}^2)$ contributions to the $1 \rightarrow 1$ amplitude.

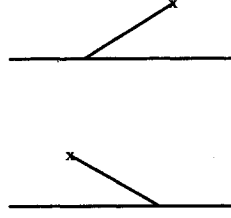


Fig. 4b) The tadpole graphs of order g_{st}^2 .

where $p_s = k_1 + k_2$, $p_t = |k_1 - k_3|$, $p_u = |k_1 - k_4|$, and

$$F(p) = \int_{-\infty}^{\infty} dk \left[\frac{(p-k)^2}{\text{sh}^2(\pi(p-k)/2)} \right] \frac{k}{p-k+i\epsilon \text{sgn}(k)}. \quad (116)$$

Using

$$\frac{1}{x-i\epsilon} = \mathcal{P}\frac{1}{x} + \pi i \delta(x) \quad (117)$$

we find $F(p) = -i\frac{4p}{\pi} - \frac{8}{3\pi}$. The total amplitude thus becomes

$$S(E_{1,2}; E_{3,4}) = -\frac{g_{st}^2}{2} \delta(E_1 + E_2 - E_3 - E_4) \prod_{j=1}^4 E_j (E_1 + E_2 + |E_1 - E_3| + |E_1 - E_4| - 2i). \quad (118)$$

Upon the Euclidean continuation $E_j \rightarrow i|q_j|$, and inclusion of the external leg factors, this precisely agrees with the non-relativistic fermion calculation of the 4-puncture correlator (56).

Now we give an example of one-loop calculation. Following ref. 41, we calculate the $\mathcal{O}(g_{st}^2)$ correction to the two-point function ($1 \rightarrow 1$ amplitude). Like the tree-level 4-point function, this is given by second-order perturbation theory. The contribution to $T(k, k)$ from the one-loop graphs is

$$\frac{18g_{st}^2 k^2}{(24\pi)^2} \int_0^{\infty} dk_1 dk_2 k_1 k_2 \left(\frac{f^2(k_1 + k_2 - k)}{k - k_1 - k_2 + i\epsilon} - \frac{f^2(k_1 + k_2 + k)}{k + k_1 + k_2 - i\epsilon} \right) \quad (119)$$

where the first term is from the 2-particle intermediate state, and the second – from

the 4-particle intermediate state (fig. 4a). After changing variables to $s = k_1 + k_2$ and k_2 , and integrating over k_2 , this reduces to

$$\frac{3g_{st}^2 k^2}{(24\pi)^2} \int_{-\infty}^{\infty} ds s^3 \frac{f^2(k-s)}{k-s+i\epsilon \operatorname{sgn}(s)} = -\frac{g_{st}^2}{48\pi} k^2 \left(ik^3 + 2k^2 + \frac{8}{15} \right) \quad (120)$$

where we evaluated the integral similarly to that of eq. (116). This is not the complete answer because there are also the diagrams of fig. 4b) arising from the tadpole term H_1 . They contribute equally, giving

$$\frac{12g_{st}^2 k^2}{(24\pi)(48\pi)} \int_0^{\infty} dk_1 k_1 \frac{f(k_1)g(k_1)}{-k_1+i\epsilon} = -\frac{g_{st}^2 k^2}{384} \int_0^{\infty} dk_1 \frac{k_1^4 + 2k_1^2}{\operatorname{sh}^2(\pi k_1/2)} = -\frac{g_{st}^2}{48\pi} \times \frac{7k^2}{15} . \quad (121)$$

Adding this to the one-loop contribution (120), we find

$$T(E, E) = -\frac{g_{st}^2}{48\pi} E^2 (iE^3 + 2E^2 + 1) . \quad (122)$$

Upon the Euclidean continuation and inclusion of the external leg factors, this agrees with eq. (54).

The agreement with the string one-loop calculation gives us further confidence that the bosonic string field theory is finite and exact. This implies that the perturbation series exhibits the $(2h)!$ behavior, which is unlike the $h!$ found in the conventional field theory. This appears to be connected with the lack of translational invariance of the hamiltonian and of the scattering amplitudes.⁴¹ In general, owing to the exponential fall-off of the form-factors $f(k)$ and $g(k)$, the hamiltonian (108) is completely free of ultraviolet divergences. Thus, our innovation of taking the double-scaling limit with $\mu < 0$ has rendered the entire perturbative expansion manifestly finite. On the other hand, the end results of our calculations do not change under the continuation $\mu \rightarrow -\mu$. This suggests that the divergences for $\mu > 0$, connected with the fact that there $f(k)$ and $g(k)$ did not fall off as $k \rightarrow \infty$,⁴¹ are simply a result of using an inconvenient formalism. This is also indicated by the fact that all divergences are correctly removed by zeta-function

regularization.⁴¹ In fact, positive and negative μ are related by the transformation that interchanges the classical coordinate λ with its conjugate momentum p . Thus, the equation

$$\left(-\frac{1}{2}\frac{\partial^2}{\partial\lambda^2} - \frac{1}{2}\lambda^2\right)\psi = -\beta\mu\psi \quad (123)$$

can be written in terms of p as

$$\left(-\frac{1}{2}\frac{\partial^2}{\partial p^2} - \frac{1}{2}p^2\right)\psi = \beta\mu\psi. \quad (124)$$

The lesson is that, while for $\mu < 0$ it is convenient to regard λ as the coordinate, for $\mu > 0$ the natural coordinate is p . If we define the original non-relativistic fermion theory in terms of $\Psi(p, t)$, then the formalism of this section goes through nicely for $\mu > 0$: there are no boundaries, and the chiralities do not mix. Thus, all the calculations above can also be regarded as $\mu > 0$ calculations which are now manifestly finite. Using this approach, the tree-level 4-point amplitude was calculated in the lagrangian formalism in ref. 42. Another technique, which works efficiently for all the tree amplitudes, but is not easily generalized to the loop corrections, was introduced in ref. 44.

Despite the impressive progress in the formulation of the bosonic string field theory of the $c = 1$ quantum gravity, there are many puzzles remaining. In particular, how should we interpret the coordinate τ and the field $X(\tau)$? Initially, it was argued that τ is essentially the zero mode of the Liouville field ϕ . If so, then it is natural to think of X as the tachyon field. However, it was shown recently that τ is not locally related to the scale factor, but is instead a conjugate variable in a complicated integral transform.³⁸ This transform is necessary to turn a string field theory, which is highly non-local in ϕ -space, into a simple local hamiltonian in τ -space. Miraculously, the matrix model has automatically provided the coordinates in which the theory looks the simplest. However, it does not appear possible to interpret X simply as the tachyon. Recall that, in addition to the tachyon, the theory has a discrete infinity of other non-field degrees of freedom. It appears that the theory has soaked up the discrete degrees of freedom together with the tachyon into a single massless field X . The form of the matrix model operators for the

discrete observables, eq. (76), indeed shows that they are mixed in the X -field together with the tachyon. Clearly, we need to attain a much better understanding of the precise string theoretic meaning of the bosonized field theory of the matrix model.

8. COMPACT TARGET SPACE AND DUALITY.

In this section we consider the discretized formulation of the sum over surfaces embedded in a circle of radius R .⁶ We will adopt the basic definition (5), with the new condition

$$G(x) = G(x + 2\pi R) \quad (125)$$

that ensures the periodicity around the circle. In this model we encounter a new set of physical issues related to the R -dependence of the sum. The $c = 1$ conformal field theory is symmetric under the transformation⁴⁵

$$R \rightarrow \frac{\alpha'}{R}; \quad g_{st} \rightarrow g_{st} \frac{\sqrt{\alpha'}}{R} \quad . \quad (126)$$

The change of g_{st} , equivalent to a constant shift of the dilaton background, is necessary to preserve the invariance to all orders of the genus expansion. A symmetry of the matter system, such as (126), is expected to survive the coupling to gravity. In this section we will find, however, that in the discretized formulation the duality is generally broken, and can only be reinstated by a careful fine-tuning. This breaking of duality is due not to the coupling to gravity, but to the introduction of the explicit lattice cut-off. Indeed, a generic lattice formulation of the compact $c = 1$ model on a fixed geometry has no dual symmetry. The phase transition that separates the small R (high temperature) phase from the large R (low temperature) phase is well-known in statistical mechanics as the Kosterlitz-Thouless transition.⁴⁶ Physically, it is due to condensation of vortices, configurations that are ignored in the “naive” continuum limit, but *are* included in the continuum limit of a generic lattice theory. The vortices are irrelevant for $R > R_{KT}$, but they condense and change the behavior of the theory for $R < R_{KT}$.

In this section we will demonstrate the lack of duality of the discretized sum over surfaces. We will also show that, if the lattice sum is modified to exclude the vortices, then the partition function is explicitly dual. These general results will be of interest when, in the next section, we consider the specific discretized sum generated by the matrix model. We will be able to isolate the effects of vortices in the matrix model, and will calculate the dual partition function in the vortex-free continuum limit.

In the conformal field theory the target space duality (126) is generated by the dual transformation on the world sheet⁴⁷

$$\partial_a X \rightarrow \epsilon_{ab} \partial_b X . \quad (127)$$

Its lattice analogue is the transformation from the lattice Λ to the dual lattice $\tilde{\Lambda}$. Recall that, when we sum over lattices Λ , we define the target space variables x_i at the centers of the faces of Λ , *i.e.* at the vertices of $\tilde{\Lambda}$. We will carry out a transformation, after which the new variables, p_I , are defined at the vertices of Λ .

First, to each directed link $\langle ij \rangle$ of $\tilde{\Lambda}$ associate the link $\langle IJ \rangle$ of Λ which intersects it, directed so that the cross product $\vec{i}\vec{j} \times \vec{I}\vec{J}$ points out of the surface (fig. 1). Now we define $D_{ij} = D_{IJ} = x_i - x_j$. Let us change variables in the integral (5) so that the integral is over D_{IJ} instead of x_i . There are E links but, apart from the zero mode, there are only $V - 1$ independent x 's.* Therefore, the D_{IJ} 's are not all independent but must satisfy $F - 1 + 2h$ constraints. The constraint associated with each face is that $\sum_{\langle ij \rangle} D_{ij} = 0$, where the sum runs over the directed boundary of a face of $\tilde{\Lambda}$. This is equivalent to the condition that the sum of the D_{IJ} 's emerging from each vertex of Λ must vanish, $\sum_J D_{IJ} = 0$. Similarly, for each independent non-contractable loop on $\tilde{\Lambda}$, of which there are $2h$, we find $\sum_{loop} D_{ij} = 0$. In terms of λ this means that $\sum_{\langle IJ \rangle} \epsilon_{IJ}^a D_{IJ} = 0$, where the symbol ϵ_{IJ}^a is non-zero only if the link IJ intersects the specially chosen non-contractable loop a on $\tilde{\Lambda}$. We direct the loop a , and define $\epsilon_{IJ}^a = 1$ if $\vec{I}\vec{J} \times \vec{a}$ points into the surface, and -1 if it points out.

* Here V , E and F are, respectively, the numbers of vertices, edges and faces of $\tilde{\Lambda}$, and h is the genus.

The theorem of Euler $V - E + F = 2 - 2h$, insures that the net number of independent variables remains unchanged. Introducing Lagrange multipliers p_I for each face of $\tilde{\Lambda}$ and l_a for each non-contractable loop, the integral over the variables x_i , for each discretization of the surface, can be replaced with

$$\begin{aligned} & \int \prod_{\langle IJ \rangle} dD_{IJ} G(D_{IJ}) \int \prod_I dp_I \exp \left[i p_I \sum_J D_{IJ} \right] \int \prod_{a=1}^{2h} dl_a \exp \left[i l_a \sum_{\langle IJ \rangle} \epsilon_{IJ}^a D_{IJ} \right] = \\ & \int \prod_I dp_I \prod_{a=1}^{2h} dl_a \prod_{\langle IJ \rangle} \tilde{G}(p_I - p_J + l_a \epsilon_{IJ}^a) \end{aligned} \quad (128)$$

where $\tilde{G}$ is the Fourier transform of G . We see that, after the transformation (128), the Lagrange multipliers p_I assume the role of new integration variables residing at the vertices of Λ . On surfaces of genus > 0 there are $2h$ additional l -integrations. Geometrically speaking, one introduces a cut on the surface along each canonically chosen non-contractable loop, so that the values of p undergo a discontinuity of l_a across the a -th cut, and subsequently integrates over l_a . On a sphere these additional integrations do not arise, and one simply includes a factor $\tilde{G}(p_I - p_J)$ for each link of the lattice Λ .

The discussion above literally applies to the string on a real line. When G is defined on a circle of radius R , then the dual variables l_a and p_I assume discrete values $\frac{n}{R}$. The integrals over p_I and l_a in eq. (128) are replaced by discrete sums, and the partition function, when expressed in terms of the dual variables, bears little resemblance to the original expression. In fact, the new variables p_I describe the embedding of the world sheet in the discretized real line with a lattice spacing $1/R$. On a sphere, where l_a do not arise, the resulting lattice sum is precisely what is needed to describe string theory with such an embedding. For higher genus, the few extra variables l_a spoil the precise equivalence, but this does not alter the essential “bulk” physics.

We have found that the transformation to the dual lattice, unlike the transformation (127) in the “naive” continuum limit, does not prove $R \rightarrow \alpha'/R$ duality. Instead, it establishes that string theory embedded in a compact dimension of radius R is essentially equivalent to string theory embedded in a discretized dimension

of lattice spacing $1/R$. To conclude that R is equivalent to $1/R$ is just as counter-intuitive as to argue that theories with very small lattice spacing in target space have the same physics as theories with enormous lattice spacing. In section 11 we will directly study the theory with the discrete target space. We will show that, if the lattice spacing is less than a critical value, then the target space lattice is “smoothed out”, but if it exceeds the critical value, then the theory drastically changes its properties. The two phases are separated by a Kosterlitz-Thouless phase transition. Equivalently, in the circle embedding the massless $c = 1$ phase for $R > R_{KT}$ is separated by the K-T transition from the disordered (massive) $c = 0$ phase for $R < R_{KT}$. The mechanism of this phase transition involves deconfinement of vortices.⁴⁶

Consider a fixed geometry $\hat{g}_{\mu\nu}$, and assume that in the continuum limit the action reduces to eq. (2). A vortex of winding number n located at the origin is described by the configuration

$$X(\theta) = nR\theta \quad (129)$$

where θ is the azimuthal angle. This configuration is singular at the origin: the values of X have a branch cut in the continuum limit. On a lattice, however, there is no singularity, and the vortex configurations are typically included in the statistical sum. If we introduce lattice spacing $\sqrt{\mu}$, then the action of a vortex is $S_n = n^2 R^2 |\ln \mu| / 4\alpha'$, and it seems that the vortices are suppressed in the continuum limit. We will now show that this expectation is only true for large enough R . Let us consider the dynamics of elementary vortices with $n = \pm 1$. Although each one is suppressed by the action, it has a large entropy: there are $\sim 1/\mu$ places on the surface where it can be found. Thus, the contribution of each vortex or antivortex to the partition function is of the order

$$\frac{1}{\mu} e^{-S_1} = \mu^{(R^2/4\alpha')-1} . \quad (130)$$

It follows that, for $R > 2\sqrt{\alpha'}$, the vortices are irrelevant in the continuum limit. On the other hand, for $R < 2\sqrt{\alpha'}$ they dominate the partition function, invalidating the “naive” continuum limit assumed in the conformal field theory.⁴⁶ In fact, in

this phase the proliferation of vortices disorders X and makes its correlations short range. The critical properties of this phase are those of pure gravity. Is there any way to salvage the “naive” continuum limit? The answer is yes, but only at the expense of fine tuning the model, so that the vortices do not appear even on the lattice. Below we find an explicit example of a model without vortices, and show that it *does* exhibit the $R \rightarrow 1/R$ duality.

First, we have to give a clear definition of a vortex on a lattice. Loosely speaking, a face I of lattice $\tilde{\Lambda}$ contains v units of vortex number if, as we follow the boundary of I , the coordinate x wraps around the target space circle v times. This does not quite define the vortex number because x is only known at a few discrete points along the boundary and cannot be followed continuously. In order to define vortex number on a lattice, it is convenient to adopt the Villain link factor⁴⁸

$$G(x) = \sum_{m=-\infty}^{\infty} e^{-\frac{1}{2}(x+2\pi mR)^2} \quad (131)$$

where the sum over m renders G periodic under $x \rightarrow x + 2\pi R$. Now, eq. (5) becomes

$$\mathcal{Z}(g_0, \kappa) = \sum_h g_0^{2(h-1)} \sum_{\Lambda} \kappa^{\text{Area}} \prod_{i=1}^V \int_0^{2\pi R} \frac{dx_i}{\sqrt{2\pi}} \prod_{\langle ij \rangle} \sum_{m_{ij}=-\infty}^{\infty} e^{-\frac{1}{2}(x_i - x_j + 2\pi m_{ij}R)^2}. \quad (132)$$

In this model the number of vortices inside a face I can be defined as

$$v_I = \sum_{\partial I} m_{ij}, \quad (133)$$

where ∂I is the directed boundary of I . After the dual transformation, one obtains a discrete sum with

$$\begin{aligned} \tilde{G}(p_I - p_J + l_a \epsilon_{IJ}^a) &= e^{-\frac{1}{2}(p_I - p_J + l_a \epsilon_{IJ}^a)^2} \\ p_I &= \frac{n_I}{R}, \quad l_a = \frac{n_a}{R} \end{aligned} \quad (134)$$

This sum is closely connected to the sum over all possible vortex numbers inside each face of $\tilde{\Lambda}$.⁴⁸

We shall now change the definition of the partition function (132) to exclude the vortex configurations. It is helpful to think of m_{ij} as link gauge fields. The vortex numbers v_I defined in eq. (133) are then analogous to field strengths. If there are no vortices, the field strength is zero everywhere. However, we still need to sum over all possible windings around non-contractable loops on the lattice $\tilde{\Lambda}$, $l_A = \sum_{loop\ a} m_{ij}$. Thus, the space of m_{ij} we need to sum over is

$$m_{ij} = \epsilon_{ij}^A l_A + m_i - m_j, \quad (135)$$

where m_i range over all integers and play the role of gauge transformations. ϵ_{ij}^A is defined above and is non-zero only for the links $\langle ij \rangle$ intersecting the specially chosen non-trivial loop A on the lattice Λ . Thus, l_A is the winding number for the non-trivial cycle a on $\tilde{\Lambda}$ which intersects the loop A . Summing over m_{ij} from eq. (135) only, the partition function becomes

$$\mathcal{Z}(g_0, \kappa) = 2\pi R \sum_h g_0^{2(h-1)} \sum_{\Lambda} \kappa^V \prod_{i=1}^{V-1} \int_{-\infty}^{\infty} \frac{dx_i}{\sqrt{2\pi}} \sum_{l_A=-\infty}^{\infty} \prod_{\langle ij \rangle} e^{-\frac{1}{2}(x_i - x_j + 2\pi l_A \epsilon_{ij}^A)^2}, \quad (136)$$

where the factor of $2\pi R$ arises from integration over the zero mode of x . After transforming to the dual lattice we obtain

$$\mathcal{Z}(g_0, \kappa) = \frac{1}{R} \sum_h \left(\frac{g_0}{\sqrt{2\pi} R} \right)^{2h-2} \sum_{\Lambda} \kappa^V \prod_{I=1}^{F-1} \int_{-\infty}^{\infty} \frac{dp_I}{\sqrt{2\pi}} \sum_{l_a=-\infty}^{\infty} \prod_{\langle IJ \rangle} e^{-\frac{1}{2}(p_I - p_J + l_a \epsilon_{IJ}^a / R)^2}. \quad (137)$$

Thus, after elimination of the vortices, the transformation to the dual lattice clearly exhibits duality under

$$\sqrt{2\pi} R \rightarrow \frac{1}{\sqrt{2\pi} R}, \quad g_0 \rightarrow \frac{g_0}{\sqrt{2\pi} R}. \quad (138)$$

Remarkably, the duality is manifest even before the continuum limit is taken. In the next section we consider the matrix model for random surfaces embedded in a circle. We will find a clear separation between the vortex contributions, and the “naive” vortex-free continuum sum. This will allow us to calculate explicitly the dual partition function of the vortex-free model.

9. MATRIX QUANTUM MECHANICS AND THE CIRCLE EMBEDDING.

The Euclidean matrix quantum mechanics is easily modified to simulate the sum over discretized random surfaces embedded in a circle of radius R : we simply define the matrix variable $\Phi(x)$ on a circle of radius R , *i.e.* $\Phi(x + 2\pi R) = \Phi(x)$. The path integral becomes⁶

$$Z = \int D^{N^2} \Phi(x) \exp \left[-\beta \int_0^{2\pi R} dx \operatorname{Tr} \left(\frac{1}{2} \left(\frac{\partial \Phi}{\partial x} \right)^2 + U(\Phi) \right) \right] . \quad (139)$$

The periodic one-dimensional propagator

$$G(x_i - x_j) = \sum_{m=-\infty}^{\infty} e^{-|x_i - x_j + 2\pi m R|} \quad (140)$$

gives the weight for each link in the random surface interpretation. In terms of the hamiltonian (19), eq. (139) is simply a path integral representation for the partition function

$$Z_R = \operatorname{Tr} e^{-2\pi R \beta H} , \quad (141)$$

so that $2\pi R$ plays the role of the inverse temperature. The problem of calculating the finite temperature partition function seems to be drastically more complicated than the zero temperature problem, where only the ground state energy was relevant. Now we have to know all the energies and degeneracies of states in arbitrarily high representations of $SU(N)$. We also have to explain the sudden enormous jump in the number of degrees of freedom as we slightly increase the temperature from zero. In view of the discussion in the last section, the reader should not be surprised if we claim that the new degrees of freedom, incorporated in the non-trivial representations of $SU(N)$, are due to the K-T vortices. Since the vortices are dynamically suppressed for $R > R_{KT}$, the problem of calculating Z_R is not as complicated as it may seem.

Let us begin by calculating the contribution to Z_R from the wave functions in the trivial representation of $SU(N)$. Since the singlet spectrum is that of N non-interacting fermions moving in the potential U , its contribution to the partition function can be calculated with the standard methods of statistical mechanics.

Instead of working with a fixed number of fermions N , we will take the well-known route of introducing a chemical potential μ_F adjusted so that

$$\kappa^2 = \frac{N}{\beta} = \int_0^\infty \rho(\epsilon) \frac{1}{1 + e^{2\pi R\beta(\epsilon - \mu_F)}} d\epsilon . \quad (142)$$

In the thermodynamic limit, $N \rightarrow \infty$, the free energy satisfies

$$\frac{\partial F}{\partial N} = \mu_F . \quad (143)$$

Shifting the variables, $\mu = \mu_c - \mu_F$, $e = \mu_c - \epsilon$, we can write the equations for the singular part of F as

$$\begin{aligned} \frac{\partial \Delta}{\partial \mu} &= \pi \tilde{\rho}(\mu) = \pi \int_{-\infty}^\infty de \rho(e) \frac{\partial}{\partial \mu} \frac{1}{1 + e^{2\pi R\beta(\mu - e)}} \\ \frac{\partial F}{\partial \Delta} &= \frac{1}{\pi} \beta \mu , \end{aligned} \quad (144)$$

to emphasize resemblance with eqs. (27) and (28) in the $R = \infty$ case. Indeed, we will calculate $F(\Delta)$ using the same method. The change introduced by a finite R only affects the first of the equations, determining $\mu(\Delta)$. We can think of $\tilde{\rho}(\mu)$ as the temperature-modified density of states. Differentiating it, we find

$$\frac{1}{\beta} \frac{\partial \tilde{\rho}}{\partial \mu} = \int_{-\infty}^\infty de \frac{1}{\beta} \frac{\partial \rho}{\partial e} \frac{\pi R \beta}{2 \cosh^2[\pi R \beta(\mu - e)]} . \quad (145)$$

Now, substituting the integral representation (35) and performing the integral over e , we find

$$\frac{1}{\beta} \frac{\partial \tilde{\rho}}{\partial \mu} = \frac{1}{\pi} \text{Im} \int_0^\infty dT e^{-i\beta\mu T} \frac{T/2}{\sinh(T/2)} \frac{T/2R}{\sinh(T/2R)} , \quad (146)$$

up to terms $\mathcal{O}(e^{-\beta\mu})$ and $\mathcal{O}(e^{-\beta\mu 2\pi R})$ which are invisible in the large- β asymptotic expansion. Thus, this integral representation should only be regarded as the gener-

ator of the correct expansion in powers of g_{st} . Integrating this equation and fixing the integration constant to agree with the WKB expansion, we find

$$\frac{\partial \Delta}{\partial \mu} = \text{Re} \int_{\mu}^{\infty} \frac{dt}{t} e^{-it} \frac{t/2\beta\mu}{\sinh t/2\beta\mu} \frac{t/2\beta\mu R}{\sinh(t/2\beta\mu R)} . \quad (147)$$

This relation has a remarkable duality symmetry under

$$R \rightarrow \frac{1}{R}, \quad \beta \rightarrow R\beta . \quad (148)$$

This is precisely the kind of duality expected in the vortex-free continuum limit. This strongly suggests that, in discarding the contributions of the non-singlet states to Z_R , we have suppressed the vortices. Later on we will offer additional arguments to support this claim.

To demonstrate the duality in the genus expansion, consider the asymptotic expansion of eq. (147)

$$\begin{aligned} \frac{\partial \Delta}{\partial \mu} &= \left[-\ln \mu + \sum_{m=1}^{\infty} \left(2\beta\mu\sqrt{R} \right)^{-2m} f_m(R) \right] , \\ f_m(R) &= (2m-1)! \sum_{k=0}^m |2^{2k} - 2| |2^{2(m-k)} - 2| \frac{|B_{2k}| |B_{2(m-k)}|}{(2k)! [2(m-k)]!} R^{m-2k} . \end{aligned} \quad (149)$$

Note that the functions $f_h(R)$ are manifestly dual: $f_1 = \frac{1}{6}(R+1/R)$, $f_2 = \frac{1}{60}(7R^2 + 10 + 7R^{-2})$, and so forth. Solving for $\mu(\Delta)$ and integrating eq. (144) we find the sum over connected surfaces $\mathcal{Z} = -2\pi R\beta F$:

$$\mathcal{Z} = \frac{1}{4} \left\{ (2\beta\mu_0\sqrt{R})^2 \ln \mu_0 - 2f_1(R) \ln \mu_0 + \sum_{m=1}^{\infty} \frac{f_{m+1}(R)}{m(2m+1)} (2\beta\mu_0\sqrt{R})^{-2m} \right\} . \quad (150)$$

Comparing this with eq. (39), valid in the case of infinite radius, we find that the coupling constant $1/(\beta\mu_0)$ has been replaced by

$$g_{eff}(R) = \frac{1}{\beta\mu_0\sqrt{R}} , \quad (151)$$

and that the contribution of each genus h contains the function $f_h(R)$. Since both $g_{eff}(R)$ and $f_h(R)$ are invariant under the dual transformation (148), we have

confirmed that the genus expansion of the sum over surfaces is manifestly dual. As in the critical string theory, the effective coupling constant depends on the radius. Indeed, the transformation $R \rightarrow 1/R$ keeping $\beta\mu_0$ fixed is not a symmetry, as it interchanges a weakly and a strongly coupled theory.

From eq. (150) we find that the sum over toroidal surfaces is $-\frac{1}{12}(R+1/R)\ln\mu_0$. This agrees with the direct calculation in the continuum, eq. (72). The extra factor of 2 in the matrix model is due to the doubling of free energy for models with symmetric potentials. To get the basic sum over triangulated surfaces, we have to divide eq. (150) by 2, which then gives perfect agreement with Liouville theory. This provides additional evidence that the singlet free energy in the matrix model evaluates the sum over surfaces in the vortex-free continuum limit. To see this more explicitly, we have to consider the non-singlet corrections to the free energy and show that, like the vortices, they are irrelevant for $R > R_{KT}$.

This problem was discussed in ref. 12, where it was argued that the total partition function can be factorized as

$$\text{Tr} e^{-2\pi R\beta H} = \text{Tr}_{\text{singlet}} e^{-2\pi R\beta H} \left(\sum_n D_n e^{-2\pi R\beta\delta E_n} \right) \quad (152)$$

where δE_n is the energy gap between the ground state and the lowest state in the n th representation, and D_n is the degeneracy factor. Let us discuss the leading correction to the free energy coming from the adjoint representation. Although the degeneracy factor $D_{adj} = N^2 - 1$ diverges as $N \rightarrow \infty$, an estimate of ref. 12 shows that the energy gap also diverges in the continuum limit,

$$\beta\delta E_{adj} = c|\ln\mu| . \quad (153)$$

As a result, the correction

$$D_{adj} e^{-2\pi R\beta\delta E_{adj}} \sim N^2 \mu^{2\pi Rc} \quad (154)$$

is negligible for $R > \frac{1}{\pi c}$ in the continuum limit where $N\mu$ is kept constant. Taking

the logarithm of eq. (152), we find that the total free energy is

$$F = F_{\text{singlet}} + \mathcal{O}\left(N^2 \mu^{2\pi R c}\right) + \dots \quad (155)$$

Thus, on the one hand, the higher representations are enhanced by enormous degeneracy factors; on the other hand, they are suppressed by energy gaps which diverge logarithmically in the continuum limit. This struggle of entropy with energy is precisely of the same type as occurs in the physics of the Kosterlitz-Thouless vortices, as discussed above. In fact, we will now argue that the leading non-singlet correction in eq. (155) is associated with a single vortex-antivortex pair. We perturb the action of the continuum theory with the operator which creates elementary vortices and anti-vortices,

$$O_v = \int d^2\sigma \sqrt{\hat{g}} \cos\left[\frac{R}{\alpha'}(X_L - X_R)\right] \quad (156)$$

where $X_{L,R}$ are the chiral components of the scalar field X . In string theory language, this is the sum of vertex operators which create states of winding number 1 and -1 . The insertion of such an operator on a surface creates an endpoint of a cut in the values of X . The conformal weight of O_v is $h = \bar{h} = \frac{R^2}{4\alpha'}$; therefore, O_v becomes relevant for $R < 2\sqrt{\alpha'}$, causing a phase transition at $R_{KT} = 2\sqrt{\alpha'}^{49}$. The phase transition is connected with the instability towards creation of cuts in the values of X , i.e., with vortex condensation.

This picture of the Kosterlitz-Thouless phase transition can be easily adapted for coupling to two-dimensional quantum gravity. The sum over genus zero surfaces coupled to the periodic scalar reduces to the path integral

$$\mathcal{Z} = N^2 \int [DX][D\phi] \exp\left(-S_L - S_0 - k \int d^2\sigma \sqrt{\hat{g}} e^{\beta\phi} \cos\left[\frac{R}{\alpha'}(X_L - X_R)\right]\right) \quad (157)$$

where S_L is the Liouville gravity action for $c = 1$, and $\beta = -2 + \frac{R}{\sqrt{\alpha'}}$ so as to make the conformal weight of the perturbing operator $(1,1)$. The gravitational dimension of the perturbation is then $\frac{R}{2\sqrt{\alpha'}} - 1$. For small values of the cut-off μ ,

and for $R > 2\sqrt{\alpha'}$,

$$\frac{\mathcal{Z}}{N^2} = -\mu^2 R \ln \mu + \mathcal{O}\left(k^2 \mu^{R/\sqrt{\alpha'}}\right) + \mathcal{O}\left(k^4 \mu^{(2R/\sqrt{\alpha'})-2}\right) + \dots, \quad (158)$$

which demonstrates that the Kosterlitz-Thouless transition occurs at $R_c = 2\sqrt{\alpha'}$, the same value as in a fixed gravitational background. The term $\mathcal{O}(k^2)$ in $\mathcal{Z}$ originates from the 2-point function of dressed winding mode vertex operators, *i.e.* from surfaces with one cut. This leading correction has the same form as eq. (155) found in the context of matrix quantum mechanics. The further corrections, coming from additional vortex-antivortex pairs, are even more suppressed for $R > R_{KT}$ in the continuum limit.

For $R < R_{KT}$, however, the expansion in k diverges badly, indicating an instability with respect to condensation of vortices. In this phase we expect the field X to become massive, and the continuum limit of the theory should behave as pure gravity. In the matrix model this is easy to show for very small R . The argument is standard: at very high temperature a $d + 1$ -dimensional theory reduces to d -dimensional theory because the compact Euclidean dimension of length $2\pi R = 1/T$ becomes so small that variation of fields along it can be ignored. Thus, for small R , the matrix quantum mechanics (139) reduces to the integral over a matrix (zero-dimensional theory)⁵⁰

$$\int D^{N^2} \Phi \exp[-2\pi R \beta \text{Tr} U(\Phi)] , \quad (159)$$

well-known to describe pure gravity.

We have accumulated a considerable amount of evidence that the non-singlet wave functions of the matrix quantum mechanics implement the physical effects of the Kosterlitz-Thouless vortices. For $R > R_{KT}$ their corrections to the partition function are negligible, and the singlet free energy (150) gives the sum over surfaces. For $R < R_{KT}$ the non-singlets dominate the partition function, and the matrix model no longer describes $c = 1$ conformal field theory coupled to gravity. In this phase the singlet free energy gives the sum over surfaces in the vortex-free “naive” continuum limit. As we showed in section 8, this continuum limit can be

achieved at the expense of a careful fine tuning of the lattice model. In the matrix model the equivalent operation turns out to be simple: just discard the non-singlet states. We conclude, therefore, that for any R the singlet sector of the matrix quantum mechanics describes $c = 1$ conformal field theory coupled to gravity. This remarkable fact means that, like the $R = \infty$ theory, the compact theory is exactly soluble in terms of free fermions. In the next section, we will use this to calculate compact correlation functions exactly to all orders in the genus expansion.

10. CORRELATION FUNCTIONS FOR FINITE RADIUS.

In this section we evaluate the correlation functions of the compact $c = 1$ model coupled to gravity using the formalism of free non-relativistic fermions at finite temperature.¹³ We have already used this approach to find the sum over surfaces with no insertions, eq. (150). The calculation of correlation functions has a very simple relation to the zero-temperature ($R = \infty$) calculation described in section 4. We again consider general operators of type (46), which reduce to fermion bilinears. Thus, the goal is to calculate the generating function $G(q_1, \lambda_1; \dots; q_n, \lambda_n)$ of eq. (48) in the compact case.

The free fermions are now described by the thermal Euclidean second-quantized action

$$S = \int_{-\infty}^{\infty} d\lambda \int_0^{2\pi R} dx \hat{\psi}^\dagger \left(-\frac{d}{dx} + \frac{d^2}{d\lambda^2} + \frac{\lambda^2}{4} + \beta\mu \right) \hat{\psi} , \quad (160)$$

and the fermion field satisfies the antiperiodic boundary conditions $\hat{\psi}(\lambda, x + 2\pi R) = -\hat{\psi}(\lambda, x)$. Compactifying the Euclidean time is a standard device for describing the theory at a finite temperature $T = 1/(2\pi R)$. The thermal vacuum satisfies

$$\langle \beta\mu | a_c^\dagger(\nu) a_{c'}(\nu) | \beta\mu \rangle = \delta_{cc'} f(\nu) \equiv \delta_{cc'} \frac{1}{e^{(\beta\mu - \nu)/T} + 1} \quad (161)$$

where $f(\nu)$ is the Fermi function and μ is the Fermi level. The thermal Euclidean Green function may be obtained by replacing $\theta(\nu - \beta\mu)$ by $f(\nu)$ in the $T = 0$ Green

function of eq. (49),

$$\begin{aligned}
 S^E(x_1, \lambda_1; x_2, \lambda_2) &= \int_{-\infty}^{\infty} d\nu e^{-(\nu - \beta\mu)\Delta x} \{ \theta(\Delta x) f(\nu) - \theta(-\Delta x)(1 - f(\nu)) \} \times \\
 \psi^\epsilon(\nu, \lambda_1) \psi^\epsilon(\nu, \lambda_2) &= \frac{i}{2\pi R} \sum_{\omega_n} e^{-i\omega_n \Delta x} \int_0^{sgn(\omega_n)\infty} ds e^{-s\omega_n + i\beta\mu s} \langle \lambda_1 | e^{2isH} | \lambda_2 \rangle.
 \end{aligned} \tag{162}$$

The only change compared to the $R = \infty$ formula (49) is that the allowed frequencies become discrete, $\omega_n = (n + \frac{1}{2})/R$, where n are integers. This ensures the antiperiodicity of S_E around the compact direction.

As in the zero-temperature case, the calculation of eq. (48) reduces to a sum over one-loop diagrams. Since the frequencies are now discrete, the single integration over the loop momentum is replaced by a sum. This is the only change introduced by the finite temperature. As we show below, its effects are easy to take into account.

First of all, each external momentum is quantized, $q_i = n_i/R$. As a result, the momentum conserving delta function $\delta(\sum q_i)$ occurring in the non-compact case is replaced by R times the Kroenecker function $\delta_{\sum q_i}$. Consider a diagram which corresponds to a given ordering of the external momentum insertions, which we denote as $q_1, q_2, \dots, q_n$. Factoring out the delta-function, the contribution of this diagram is

$$\begin{aligned}
 \frac{i^n}{R} \sum_{\omega_n} \int_0^{sgn(\omega_n)\infty} d\alpha_1 e^{-\alpha_1 \omega_n + i\beta\mu \alpha_1} \langle \lambda_n | e^{2i\alpha_1 H} | \lambda_1 \rangle \int_0^{sgn(\omega_n + q_1)\infty} d\alpha_2 e^{-\alpha_2(\omega_n + q_1) + i\beta\mu \alpha_2} \times \\
 \langle \lambda_1 | e^{2i\alpha_2 H} | \lambda_2 \rangle \dots \int_0^{sgn(\omega_n + \sum_{i=1}^{n-1} q_i)\infty} d\alpha_n e^{-\alpha_n(\omega_n + \sum_{i=1}^{n-1} q_i) + i\beta\mu \alpha_n} \langle \lambda_{n-1} | e^{2i\alpha_n H} | \lambda_n \rangle
 \end{aligned} \tag{163}$$

It is necessary to break the sum over ω_n into parts where the sgn functions in the exponents are constant. Thus, we form $p_k = -\sum_{i=1}^k q_i$ and put the p 's in the

increasing order. If p and p' are two consecutive p 's, then the summation that needs to be performed is, defining $\xi = \sum_1^n \alpha_i$,

$$\frac{1}{R} \sum_{\omega_n=p+\frac{1}{2R}}^{p'-\frac{1}{2R}} e^{-\omega_n \xi} = \frac{e^{-p\xi} - e^{-p'\xi}}{\xi} \frac{\xi/2R}{\text{sh}(\xi/2R)}. \quad (164)$$

This factor multiplies the integrand of an n -fold α_i -integral. Comparing this with the $R = \infty$ case, we find that there the loop momentum integration between p and p' gives the factor $(e^{-p\xi} - e^{-p'\xi})/\xi$ multiplying the same n -fold integrand. It follows that, in every single contribution to the final answer, the compactification of the target space simply inserts the extra factor $\frac{\xi/2R}{\text{sh}(\xi/2R)}$ into the integrand. Thus, from eq. (51) we arrive at

$$\begin{aligned} \frac{1}{\beta} \frac{\partial}{\partial \mu} G_R(q_1, \lambda_1; \dots; q_n, \lambda_n) &= i^{n+1} R \delta_{\sum q_i} \sum_{\sigma \in \Sigma_n} \int_{-\infty}^{\infty} d\xi e^{i\beta \mu \xi} \frac{\xi/2R}{\text{sh}(\xi/2R)} \int_0^{\epsilon_1 \infty} ds_1 \dots \\ &\int_0^{\epsilon_{n-1} \infty} ds_{n-1} e^{-s_1 Q_1^\sigma - \dots - s_{n-1} Q_{n-1}^\sigma} \langle \lambda_{\sigma(1)} | e^{2is_1 H} | \lambda_{\sigma(2)} \rangle \dots \langle \lambda_{\sigma(n)} | e^{2i(\xi - \sum_1^{n-1} s_i) H} | \lambda_{\sigma(1)} \rangle \end{aligned} \quad (165)$$

This formula allows us to calculate correlation functions of any set of operators of type (46) in the compact $c = 1$ model. The integral representation of any such correlator is obtained from that in the non-compact case by insertion of the factor $\frac{\xi/2R}{\text{sh}(\xi/2R)}$ into the integrand, and by replacement $\delta(\sum q_i) \rightarrow R \delta_{\sum q_i}$. A special case of this rule is evident in passing from the $R = \infty$ equation (25) to the finite R equation (146).

In general, comparing eqs. (165) and (51), we find that, if

$$\langle \mathcal{O}_1(q_1) \dots \mathcal{O}_n(q_n) \rangle_{R=\infty} = \delta(\sum q_i) F(q_1, \dots, q_n; \mu, \beta)$$

then

$$\begin{aligned} \langle \mathcal{O}_1(q_1) \dots \mathcal{O}_n(q_n) \rangle_R &= R \delta_{\sum q_i} F_R(q_1, \dots, q_n; \mu, \beta), \\ F_R(q_1, \dots, q_n; \mu, \beta) &= \frac{\frac{1}{2R\beta} \frac{\partial}{\partial \mu}}{\sin(\frac{1}{2R\beta} \frac{\partial}{\partial \mu})} F(q_1, \dots, q_n; \mu, \beta). \end{aligned} \quad (166)$$

This remarkable relation exists because the entire dependence on μ is through the factor $e^{i\beta \mu \xi}$ in the integrands of eqs. (51) and (165). Therefore, a function $f(\xi)$

can be introduced into the integrand simply by acting on the integral with operator $f(-\frac{i}{\beta} \frac{\partial}{\partial \mu})$.

Since μ has the interpretation of the renormalized cosmological constant, in each sum over surfaces of fixed area the compactification simply introduces the non-perturbative factor $\frac{A/2R\beta}{\sin(A/2R\beta)}$. The remarkable feature of this modification, and, equivalently, of the operator in eq. (166), is that they are entirely independent of which operators are inserted into the surface. To demonstrate the power of eq. (166), we will use the non-compact correlators of eqs. (54)-(56) to develop the corresponding genus expansions in the compact case with virtually no extra work. To this end, we expand the operator in eq. (166) as

$$\frac{\frac{1}{2R\beta} \frac{\partial}{\partial \mu}}{\sin(\frac{1}{2R\beta} \frac{\partial}{\partial \mu})} = 1 + \sum_{k=1}^{\infty} \frac{(1 - 2^{-2k+1})|B_{2k}|}{(\beta R)^{2k}(2k)!} \left(\frac{\partial}{\partial \mu} \right)^{2k}. \quad (167)$$

Substituting this expansion into eq. (166), we determine the finite R genus g correlators in terms of the $R = \infty$ genus $\leq g$ correlators,

$$\begin{aligned} F_R^{g=0}(q_i; \mu) &= F^{g=0}(q_i; \mu), \\ F_R^{g=1}(q_i; \mu) &= F^{g=1}(q_i; \mu) + \frac{1}{24R^2} \frac{\partial^2}{\partial \mu^2} F^{g=0}(q_i; \mu), \\ F_R^{g=2}(q_i; \mu) &= F^{g=2}(q_i; \mu) + \frac{1}{24R^2} \frac{\partial^2}{\partial \mu^2} F^{g=1}(q_i; \mu) + \frac{7}{5760R^4} \frac{\partial^4}{\partial \mu^4} F^{g=0}(q_i; \mu), \text{ etc.} \end{aligned} \quad (168)$$

In particular, apart from the discretization of momenta and the change in the overall δ -function, all spherical correlation functions are not affected by the compactification. This was, of course, expected in the Liouville approach. The advantage of the matrix model formalism is that the entire genus expansion of any correlator can be calculated. Surprisingly, as we have demonstrated, the compact case is hardly any more difficult than the non-compact case. From the point of view of the Liouville approach, this conclusion is quite remarkable. There the finite R calculations have the additional complication of having to perform $2h$ sums over the winding numbers around the non-trivial cycles of a genus h surface. Thus, eqs. (166) and (168), which are completely general, pose a new challenge for the Liouville approach.

Below we give genus expansions for some correlation functions of puncture operators $P(q)$. For the two-point function up to genus three we get

$$\begin{aligned}
\langle P(q)P(-q) \rangle = & -R[\Gamma(1-|q|)]^2 \mu^{|q|} \left[\frac{1}{|q|} - \frac{(\sqrt{R}\beta\mu)^{-2}}{24}(|q|-1) \times \right. \\
& \left\{ R(q^2 - |q| - 1) - \frac{1}{R} \right\} + \frac{(\sqrt{R}\beta\mu)^{-4}}{5760} \prod_{r=1}^3 (|q| - r) \times \\
& \left\{ R^2(3q^4 - 10|q|^3 - 5q^2 + 12|q| + 7) - 10(q^2 - |q| - 1) + \frac{7}{R^2} \right\} \\
& - \frac{(\sqrt{R}\beta\mu)^{-6}}{2903040} \prod_{r=1}^5 (|q| - r) \left\{ R^3(9q^6 - 63|q|^5 + 42q^4 + 217|q|^3 - 205|q| - 93) \right. \\
& \left. \left. - 21R(3q^4 - 10|q|^3 - 5q^2 + 12|q| + 7) + \frac{147}{R}(q^2 - |q| - 1) - \frac{93}{R^3} \right\} + \dots \right].
\end{aligned} \tag{169}$$

We see that the external leg factors are the same as in the non-compact case. The relation between the tachyon operators $T(q)$ and $P(q)$ is also the same, eq. (64). For the same reason as in the non-compact case, eq. (169) is, strictly speaking, valid only for non-integer q , where we may replace μ by μ_0 . As in the sum over surfaces, we find that the effective string coupling constant is $g_{eff} \sim \frac{1}{\sqrt{R}\beta\mu_0}$. The R -dependence of the correlator at each genus is not dual: the duality is broken by insertions of momentum into the world sheet. In the limit $q_i \rightarrow 0$ the duality is restored, and the tachyon correlators reduce to derivatives of the sum over surfaces $\mathcal{Z}(\mu_0, R)$,

$$\langle T(0)T(0) \rangle \sim \frac{\partial^2}{\partial \Delta^2} \mathcal{Z}(\mu_0, R). \tag{170}$$

The three-point function for the case $q_1, q_2 > 0$ is, up to genus two,

$$\begin{aligned}
\langle T(q_1)T(q_2)T(q_3) \rangle &= R\delta_{\sum q_i} \prod_{i=1}^3 \left(\frac{\Gamma(1-|q_i|)}{\Gamma(|q_i|)} \mu^{\frac{|q_i|}{2}} \right) \frac{1}{\beta\mu} \left[1 \right. \\
&- \frac{(\sqrt{R}\beta\mu)^{-2}}{24} (|q_3| - 1)(|q_3| - 2) \left\{ R(q_1^2 + q_2^2 - |q_3| - 1) - \frac{1}{R} \right\} + \prod_{r=1}^4 (|q_3| - r) \times \\
&\left\{ R^2 \left(3(q_1^4 + q_2^4) + 10q_1^2 q_2^2 - 10(q_1^2 + q_2^2)|q_3| - 5(q_1 - q_2)^2 + 12|q_3| + 7 \right) \right. \\
&\left. \left. - 10(q_1^2 + q_2^2 - |q_3| - 1) + \frac{7}{R^2} \right\} \frac{(\sqrt{R}\beta\mu)^{-4}}{5760} + \dots \right].
\end{aligned} \tag{171}$$

In the case of the four-point function we must consider two independent kinematic configurations which will give amplitudes that are not related by analytic continuation.¹⁹ The simplest case is when $q_1, q_2, q_3 > 0$,

$$\begin{aligned}
\langle T(q_1)T(q_2)T(q_3)T(q_4) \rangle &= -R\delta_{\sum q_i} \prod_{i=1}^4 \left(\frac{\Gamma(1-|q_i|)}{\Gamma(|q_i|)} \mu^{\frac{|q_i|}{2}} \right) (\beta\mu)^{-2} \left[(|q_4| - 1) \right. \\
&- \frac{(\sqrt{R}\beta\mu)^{-2}}{24} \prod_{r=1}^3 (|q_4| - r) \left\{ R(q_1^2 + q_2^2 + q_3^2 - |q_4| - 1) - \frac{1}{R} \right\} \\
&+ \frac{(\sqrt{R}\beta\mu)^{-4}}{5760} \prod_{r=1}^5 (|q_4| - r) \left\{ R^2 \left(3(q_1^4 + q_2^4 + q_3^4) + 10(q_1^2 q_2^2 + q_1^2 q_3^2 + q_2^2 q_3^2) \right. \right. \\
&- 10|q_4|(q_1^2 + q_2^2 + q_3^2) - 5(q_1^2 + q_2^2 + q_3^2) + 10(q_1 q_2 + q_1 q_3 + q_2 q_3) + 12|q_4| + 7) \\
&\left. \left. - 10(q_1^2 + q_2^2 + q_3^2 - |q_4| - 1) + \frac{7}{R^2} \right\} + \dots \right].
\end{aligned} \tag{172}$$

The other kinematic configuration is $q_1, q_2 > 0$, $q_3, q_4 < 0$ with $\min(|q_i|) = q_1$, $\max(|q_i|) = q_2$. This leads to a much more complicated formula, so we just give the answer out to genus one

$$\begin{aligned}
\langle T(q_1)T(q_2)T(q_3)T(q_4) \rangle &= -R\delta_{\sum q_i} \prod_{i=1}^4 \left(\frac{\Gamma(1-|q_i|)}{\Gamma(|q_i|)} \mu^{\frac{|q_i|}{2}} \right) \frac{q_2 - 1}{(\beta\mu)^2} \left[1 - \frac{(\sqrt{R}\beta\mu)^{-2}}{24} \right. \\
&\times \left\{ R \left(S^3(S - 2q_2 - 6) + q_2^2(S^2 + q_3^2 + q_4^2) + 13S^2 q_2 - 8S q_2^2 - 5q_2(q_3^2 + q_4^2) \right. \right. \\
&\left. \left. + 6S(S - 2q_2) + 6(q_3^2 + q_4^2) + 10q_2^2 + S - 2q_2 - 6 \right) - \frac{1}{R}(S - 2)(S - 3) \right\} + \dots \Big]
\end{aligned} \tag{173}$$

where we have defined $S = \frac{1}{2} \sum |q_i|$.

Using the matrix model, we have directly calculated the correlation functions of the tachyon operators for the momentum modes,

$$T(q) = \frac{1}{\beta} \int d^2\sigma \sqrt{\hat{g}} e^{iq(X+\bar{X})} e^{(-2+|q|)\phi} , \quad (174)$$

where X and $\bar{X}$ are the holomorphic and anti-holomorphic parts of the X -field. In string theory we also have the winding operators

$$\tilde{T}(\tilde{q}) = \frac{1}{\beta} \int d^2\sigma \sqrt{\hat{g}} e^{i\tilde{q}(X-\bar{X})} e^{(-2+|\tilde{q}|\phi)} , \quad (175)$$

where $\tilde{q}$ is quantized as nR . Although these operators are hard to introduce in the matrix model directly, their correlators are obtained from those of the momentum operators through the dual transformation

$$X + \bar{X} \rightarrow X - \bar{X} , \quad R \rightarrow \frac{1}{R} , \quad \beta \rightarrow \beta R. \quad (176)$$

Carrying out the dual transformation (176) on eqns. (169)-(173), we find

$$\begin{aligned} \langle \tilde{T}(\tilde{q}) \tilde{T}(-\tilde{q}) \rangle = & R \left[\frac{\Gamma(1-|\tilde{q}|)}{\Gamma(|\tilde{q}|)} \right]^2 \mu^{|\tilde{q}|} \left[\frac{1}{|\tilde{q}|} \right. \\ & \left. - \frac{(\sqrt{R}\beta\mu)^{-2}}{24} (|\tilde{q}| - 1) \left\{ \frac{1}{R} (\tilde{q}^2 - |\tilde{q}| - 1) - R \right\} + \dots \right] , \end{aligned} \quad (177)$$

etc. In general, if

$$\begin{aligned} \langle T(q_1) \dots T(q_n) \rangle_{R=\infty} = & \delta(\sum q_i) F(q_1, \dots, q_n; \mu, \beta) , \quad \text{then} \\ \langle \tilde{T}(\tilde{q}_1) \dots \tilde{T}(\tilde{q}_n) \rangle_R = & R^{n-1} \delta_{\sum \tilde{q}_i} \frac{\frac{1}{2\beta} \frac{\partial}{\partial \mu}}{\sin(\frac{1}{2\beta} \frac{\partial}{\partial \mu})} F(\tilde{q}_1, \dots, \tilde{q}_n; \mu, \beta R) . \end{aligned} \quad (178)$$

This formula was obtained by carrying out the dual transformation on the similar formula (166) for the momentum states. Since we do not have a direct matrix model calculation of the winding operator correlations, we do not yet know the expressions for correlators where both the momentum and the winding operators are present.

Is there any real time interpretation of the exact correlation functions we have calculated? In the $R = \infty$ case we found that, after the continuation to Minkowski signature, the external leg factors turn into pure phases, and the correlation functions become scattering amplitudes of the X -quanta of the Das-Jevicki field theory. We can perform the same continuation, $|q_j| \rightarrow -iE_j$, for the finite R correlation functions, expecting to obtain the same field theory at a finite temperature $T = 1/(2\pi R)$. The external leg factors again turn into phases and can be absorbed in the definition of the vertex operators, and we find the scattering amplitudes at temperature T . For instance,

$$S(E_1, E_2; E_3) = -g_{str} \delta(E_1 + E_2 - E_3) E_1 E_2 E_3 \left[1 + \frac{g_{str}^2}{24} (1 + iE_3)(2 + iE_3)(1 - iE_3 + E_1^2 + E_2^2 + (2\pi T)^2) + \dots \right] \quad (179)$$

Is the concept of finite-temperature S -matrix well-defined? In the usual theories the answer is negative, because interaction with the heat bath stops a particle before it reaches the interaction region. However, in the hamiltonian (97) the interaction is exponentially localized in a finite region of space near $\tau = 0$. Thus, the interaction with the heat bath is negligible at infinity, and the concept of S -matrix still appears to be meaningful. It would be interesting to compare the fermionic results like eq. (179) with bosonic finite temperature calculations with the hamiltonian (97).

In conclusion, we should also mention the fascinating speculation due to Witten that the $c = 1$ matrix model describes physics in the background of a two-dimensional black hole.^{51*} The black hole has finite temperature, and its Euclidean description is, perhaps, related to the compact $c = 1$ model with a definite radius R . It would be remarkable if the formulae reported in this section could be interpreted in terms of massless particle scattering off the black hole and in terms of Hawking radiation.

* For a detailed discussion, see H. Verlinde's lectures in this volume.

11. DISCRETIZED TARGET SPACE.

In this section we will study another interesting model: the theory of surfaces embedded in a discretized real line with lattice spacing ϵ . As shown in section 8, the lattice duality transformation relates it to the model of surfaces embedded in a circle of radius $1/\epsilon$. Although for genus > 0 this transformation generates $2h$ extra variables which violate the precise equivalence of the two theories, the basic physical effect – the Kosterlitz-Thouless phase transition – is of the same nature.

In fact, the model with discretized target space is very interesting in its own right. Here we can test how string theory responds to introduction of a small lattice spacing in space-time. It has been argued⁵² that in this respect string theory is very different from any field theory: if the lattice spacing is smaller than some critical value, then the theory on a lattice is *precisely* the same as the theory in the continuum. This is to be contrasted with any known field theory, where the continuum behavior may only be recovered as the lattice spacing is sent to zero. The exactly soluble $c = 1$ string theory is an ideal ground for testing this remarkable stringy effect. Below we give the exact solution of the model with the discretized target space and confirm that, for $\epsilon < \epsilon_c$, the lattice is smoothed out so that the theory is identical to the embedding in R^1 .

The matrix model representation of the partition function is now in terms of an integral over a chain of M matrices with nearest neighbor couplings⁵³

$$Z(\epsilon) = \prod_{i=1}^M \int D^{N^2} \Phi_i \exp \left[-\beta \sum_i \left(\frac{1}{2\epsilon} \text{Tr}(\Phi_{i+1} - \Phi_i)^2 + \epsilon \text{Tr} W(\Phi_i) \right) \right] . \quad (180)$$

It is easy to check that the perturbative expansion of this integral gives the statistical sum of the form

$$\lim_{M \rightarrow \infty} \frac{\ln Z(\epsilon)}{M} = \sum_h g_0^{2(h-1)} \sum_{\Lambda} \kappa^V \prod_{i=1}^{V-1} \sum_{n_i} \prod_{\langle ij \rangle} e^{-\epsilon |n_i - n_j|} , \quad (181)$$

which is precisely what is needed to describe the embedding in a discretized real line with lattice spacing ϵ . As in the matrix quantum mechanics (6), the integral (180)

can be expressed in terms of the eigenvalues of the matrices Φ_i . The only modification here is that, instead of quantum mechanics of N identical non-interacting fermions, we now find quantum mechanics with a discrete time step ϵ . Thus,

$$\lim_{M \rightarrow \infty} \frac{\ln Z(\epsilon)}{M} = \sum_{i=1}^N \ln \mu_i, \quad (182)$$

where μ_i are the N largest eigenvalues of the transfer matrix

$$\begin{aligned} \mu_i f_i(x) &= \int_{-\infty}^{\infty} dy K(x, y) f_i(y), \\ K(x, y) &= \left(\frac{\beta}{2\pi\epsilon} \right)^{1/2} \exp \left[-\frac{\beta}{2} \left\{ \frac{(x-y)^2}{\epsilon} + \epsilon(W(x) + W(y)) \right\} \right]. \end{aligned} \quad (183)$$

Although it is hard to solve this problem exactly, as usual things simplify in the continuum limit where we expect to find universal behavior controlled by the quadratic maximum of $W(x)$. Indeed, if we take

$$W(x) = -\frac{1}{2}x^2 + \mathcal{O}(x^3) \quad (184)$$

and rescale $x\sqrt{\beta} = z$, we find

$$K(z, w) = \frac{1}{\sqrt{2\pi\epsilon}} \exp \left[-\frac{(z-w)^2}{2\epsilon} + \frac{1}{4}\epsilon \left(z^2 + w^2 + \mathcal{O}\left(\frac{1}{\sqrt{\beta}}\right)(z^3 + w^3) \right) \right]. \quad (185)$$

Thus, the terms beyond the quadratic order are suppressed by powers of β and are, therefore, irrelevant. The quadratic problem can be solved exactly through finding the equivalent quantum mechanical problem with Planck constant $1/\beta'$ and hamiltonian $H(\epsilon)$ such that

$$K(x, y) = \langle x | e^{-\epsilon\beta' H(\epsilon)} | y \rangle. \quad (186)$$

Then, $\mu_i = \exp(-\epsilon\beta' e_i)$ where e_i are the N lowest eigenvalues of $H(\epsilon)$. For the

quadratic transfer matrix,

$$K(x, y) = \left(\frac{\beta}{2\pi\epsilon} \right)^{1/2} \exp \left[-\frac{\beta}{2} \left(\frac{(x-y)^2}{\epsilon} - \frac{1}{2}\epsilon(x^2 + y^2) \right) \right] , \quad (187)$$

the hamiltonian is also quadratic $H = \frac{p^2}{2} - \frac{\omega^2 x^2}{2}$. Comparing

$$\langle x | e^{-\epsilon\beta'H} | y \rangle = \left(\frac{m\omega\beta'}{2\pi \sin \omega\epsilon} \right)^{1/2} \exp \left[-\frac{m\omega\beta'}{2} ((x^2 + y^2) \cot \omega\epsilon - 2xy \sin^{-1} \omega\epsilon) \right] \quad (188)$$

with eq. (187), we find

$$\cos \epsilon\omega = 1 - \frac{1}{2}\epsilon^2 , \quad \beta' = \beta \frac{\sin(\omega\epsilon)}{\omega\epsilon} . \quad (189)$$

Thus, we have reduced the problem of random surfaces embedded in discretized target space to quantum mechanics of N fermions moving in an inverted harmonic oscillator potential, *i.e.* to random surfaces embedded in R^1 ! Eq. (189) dictates how the curvature of the potential and β' depend on ϵ . Recalling that $\alpha' = 1/\omega^2$ and $g_{st} = \frac{1}{\beta'\mu_0}$, we find that the target space scale and the coupling constant of the equivalent continuum string theory depend on ϵ . Thus, changing the lattice spacing in target space simply changes the parameters of the equivalent string theory embedded in continuous space. This effect was originally demonstrated in higher-dimensional string theory using completely different techniques.⁵² It appears to be a general property of string theory.

Intuitively, we do not expect the equivalence with the continuum string theory to be there for arbitrarily large ϵ . Indeed, eq. (189) shows that, if ϵ exceeds $\epsilon_c = 2$ then the solution for ω becomes complex, which is a sign of instability. Since large lattice spacing corresponds, in the dual language, to small radius, it is clear that the instability is with respect to condensation of vortices. Thus, for $\epsilon > \epsilon_c$, we expect the vortices to eliminate the massless field on the world sheet corresponding to the embedding dimension, so that the pure gravity results. This is easy to check for very large ϵ where the sites of the matrix chain become decoupled,

$$\lim_{M \rightarrow \infty} \frac{\ln Z(\epsilon)}{M} \rightarrow \int D^{N^2} \Phi e^{-\beta\epsilon \text{Tr } W(\Phi)}, \quad (190)$$

and we obtain the one-matrix model well-known to describe the pure gravity.

We will now give a continuum explanation of the Kosterlitz-Thouless transition from the small ϵ phase, where the lattice is irrelevant, to the large ϵ phase, where it has the dominant effect. We will replace the discretized real line by a continuous variable X with a periodic potential $V(X) = \sum_{n>0} a_n \cos(\frac{2\pi n}{\epsilon} X)$, which implements the effects of the lattice. The conformal weight of the operator $\int d^2\sigma \cos(\frac{2\pi n}{\epsilon} X)$ is $h = \frac{\pi^2 n^2 \alpha'}{\epsilon^2}$. Thus, for small ϵ all the operators perturbing the action have $h > 1$ and are irrelevant. This is the essential reason why the string does not feel a lattice in target space with spacing $< \mathcal{O}(\sqrt{\alpha'})$. The phase transition to lattice-dominated phase takes place when the perturbation with $n = 1$ becomes relevant, *i.e.*, for

$$\epsilon_c = \pi \sqrt{\alpha'} \quad (191)$$

We can compare this result with the position of the K-T transition in the matrix chain model, where we found $\epsilon_c = 2$ and $\sqrt{\alpha'}(\epsilon_c) = 2/\pi$. The agreement of these values with the continuum relation (191) strengthens our explanation of the phase transition in the matrix model. It is satisfying that we have found the matrix model confirmation of a general effect, smoothing out of a target space lattice, which applies to string theories in any dimension.

12. CONCLUSION.

I hope that I have convinced the reader that the 2-dimensional string theory is a highly non-trivial toy model for string theory in higher dimensions. Its matrix model formulation as a sum over surfaces embedded in 1 dimension is an example of a perfectly regularized generally covariant definition of the Polyakov path integral. It also turns out to be remarkably powerful, giving us the exact solution of a non-trivial string theory. Many of its physical features, such as the $R \rightarrow 1/R$ duality (and its breaking due to the vortices), smoothing out of the target space discreteness, presence of poles in the correlation functions, *etc.*, carry over to string theories in higher dimensions. Also, even though this string theory is two-dimensional, it has some interesting remnants of transverse excitations, manifested in the isolated states at integer momenta which generate the $W_{1+\infty}$ algebra. Thanks to the matrix models, we have acquired a wealth of exact information on the perturbative

properties of this string theory. Unfortunately, we still do not have a detailed understanding or interpretation of most of it in the conventional continuum formalism. Hopefully, this gap will be closed in the not too distant future.

Acknowledgements: I am grateful to D. Gross, D. Lowe, M. Bershadsky and M. Newman, in collaboration with whom much of this material was developed. I am indebted to A. Polyakov for many helpful discussions and comments. I also thank D. Lowe for reading the manuscript, and R. Wilkinson for his help in making figures. I am grateful to ICTP and the organizers of the Spring School for their hospitality.

REFERENCES

1. V. Kazakov, *Phys. Lett.* **150**, 282 (1985); J. Ambjørn, B. Durhuus, and J. Fröhlich *Nucl. Phys.* **B257**, 433 (1985), F. David, *Nucl. Phys.* **B 257**, 45 (1985); V. Kazakov, I. Kostov and A. Migdal, *Phys. Lett.* **157**, 295 (1985)
2. D. J. Gross and A. A. Migdal, *Phys. Rev. Lett.* **64** (1990) 717; M. Douglas and S. Shenker, *Nucl. Phys.* **B335** (1990) 635; E. Brezin and V. Kazakov, *Phys. Lett.* **236B** (1990) 144.
3. D. J. Gross and A. Migdal, *Phys. Rev. Lett.* **64** (1990) 717; E. Brezin, M. Douglas, V. Kazakov and S. Shenker, *Phys. Lett.* **237B** (1990) 43; C. Crnkovic, P. Ginsparg and G. Moore, *Phys. Lett.* **237B** (1990) 196.
4. M. Douglas, *Phys. Lett.* **238B** (1990) 176.
5. D. J. Gross and N. Miljkovic, *Phys. Lett.* **238B** (1990), 217; E. Brezin, V. Kazakov, and Al. B. Zamolodchikov, *Nucl. Phys.* **B338** (1990), 673; P. Ginsparg and J. Zinn-Justin, *Phys. Lett.* **240B** (1990), 333; G. Parisi, *Phys. Lett.* **238B** (1990), 209.
6. D. J. Gross and I. R. Klebanov, *Nucl. Phys.* **B344** (1990) 475.
7. S. Das, S. Naik and S. Wadia, *Mod. Phys. Lett.* **A4** (1989) 1033; J. Polchinski, *Nucl. Phys.* **B324** (1989) 123; S. Das, A. Dhar and S. Wadia, *Mod. Phys. Lett.* **A5** (1990) 799; T. Banks and J. Lykken, *Nucl. Phys.* **B331** (1990) 173; A. Tseytlin, *Int. Jour. Mod. Phys.* **A5** (1990) 1833.

8. E. Brezin, C. Itzykson, G. Parisi and J. Zuber, *Comm. Math. Phys.* **59** (1978) 35.
9. S. R. Das and A. Jevicki, *Mod. Phys. Lett.* **A5** (1990), 1639.
10. D. J. Gross and I. R. Klebanov, *Nucl. Phys.* **B352** (1991), 671.
11. A. M. Polyakov, *Phys. Lett.* **103B** (1981) 207.
12. D. J. Gross and I. R. Klebanov, *Nucl. Phys.* **B354** (1991), 459.
13. I. R. Klebanov and D. Lowe, Princeton preprint PUPT-1256, to appear in *Nucl. Phys. B*.
14. V. Knizhnik, A. Polyakov and A. Zamolodchikov, *Mod. Phys. Lett.* **A3** (1988) 819.
15. F. David, *Mod. Phys. Lett.* **A3** (1988) 1651; J. Distler and H. Kawai, *Nucl. Phys.* **B321** (1989) 509.
16. V. Kazakov and A. Migdal, *Nucl. Phys.* **B311** (1989) 171.
17. P. Marchesini and E. Onofri, *J. Math. Phys.* **21** (1980) 1103.
18. S. Shenker, Rutgers preprint RU-90-47 (1990).
19. G. Moore, Yale and Rutgers preprint YCTP-P8-91, RU-91-12.
20. D. J. Gross, I. R. Klebanov and M. J. Newman, *Nucl. Phys.* **B350** (1991), 621.
21. J. Polchinski, *Nucl. Phys.* **B346** (1990), 253.
22. T. L. Curtright and C. B. Thorn, *Phys. Rev. Lett.* **48** (1982) 1309.
23. N. Seiberg, Rutgers preprint RU-90-29.
24. J. Polchinski, Proceedings of *Strings '90*, Texas preprint UTTG-19-90 (1990).
25. S. Gupta, A. Trivedi and M. Wise, *Nucl. Phys.* **B340** (1990) 475.
26. M. Goulian and M. Li, Santa Barbara preprint UCSBTH-90-61.
27. P. Di Francesco and D. Kutasov, *Phys. Lett.* **B261** (1991) 385.
28. M. Bershadsky and I. R. Klebanov, *Phys. Rev. Lett.* **65** (1990), 3088; N. Sakai and Y. Tanii, Tokyo Inst. of Tech. preprint TIT/HEP-160 (1990).

29. M. Bershadsky and I. R. Klebanov, *Nucl. Phys.* **B360** (1991), 559.
30. B. McClain and B. D. B. Roth, *Comm. Math. Phys.* **111** (1987) 539; K. H. O'Brien and C.-I. Tan, *Phys. Rev.* **D36** (1987) 1184.
31. J. Polchinski, *Comm. Math. Phys.* **104** (1986) 37.
32. A. M. Polyakov, *Mod. Phys. Lett.* **A6** (1991) 635.
33. B. Lian and G. Zuckerman, Yale preprint YCTP-P18-91.
34. R. Dijkgraaf, E. Verlinde and H. Verlinde, *Comm. Math. Phys.* **115** (1988) 649.
35. E. Witten, paper in preparation.
36. U. H. Danielsson and D. J. Gross, Princeton preprint PUPT-1258.
37. J. Avan and A. Jevicki, Brown preprint BROWN-HET-801; D. Minic, J. Polchinski and Z. Yang, Univ. of Texas preprint UTTG-16-91.
38. G. Moore and N. Seiberg, Rutgers and Yale preprint RU-91-29, YCTP-P19-91.
39. A. Sengupta and S. Wadia, *Int. Jour. Mod. Phys.* **A6** (1991) 1961.
40. S. Coleman, *Phys. Rev.* **D11** (1975) 2088; S. Mandelstam, *Phys. Rev.* **D11** (1975) 3026.
41. K. Demeterfi, A. Jevicki and J. P. Rodrigues, Brown preprints BROWN-HET-795 and BROWN-HET-803 (1991).
42. G. Mandal, A. Sengupta and S. R. Wadia, IAS preprint IASSNS-HEP/91/8 (1991).
43. D. J. Gross and I. R. Klebanov, *Nucl. Phys.* **B359** (1991) 3.
44. J. Polchinski, University of Texas preprint UTTG-06-91.
45. P. Ginsparg and C. Vafa, *Nucl. Phys.* **B289** (1987) 414; E. Alvarez and M. Osorio, *Phys. Rev.* **D40** (1989), 1150.
46. V. L. Berezinskii, *JETP* **34** (1972) 610; M. Kosterlitz and D. Thouless, *J. Phys.* **C6** (1973) 1181.

- 47. J. Molera and B. Ovrut, *Phys. Rev.* **D40** (1989), 1146; M. J. Duff, preprint CTP-TAMU-53/89.
- 48. J. Villain, *J. Phys.* **C36** (1975) 581.
- 49. Ya. I. Kogan, *JETP Lett.* **45** (1987), 709; B. Sathiapalan, *Phys. Rev.* **D35** (1987), 3277.
- 50. Z. Yang, *Phys. Lett.* **243B** (1990) 365.
- 51. E. Witten, Inst. for Advanced Study preprint IASSNS-HEP-91/12.
- 52. I. Klebanov and L. Susskind, *Nucl. Phys.* **B309** (1988) 175
- 53. G. Parisi, *Phys. Lett.* **238B** (1990) 213.

Some Properties of (Non) Critical Strings

D. Kutasov

*Joseph Henry Laboratories,
Princeton University,
Princeton, NJ 08544.*

We review some recent developments in string theory, emphasizing the importance of vacuum instabilities, their relation to the density of states, and the role of space-time fermions in non-critical string theory. We also discuss the classical dynamics of two dimensional string theory.

1. INTRODUCTION.

In the last few years many people have been studying simple models of string theory, which correspond to string propagation in two dimensional space-time (and simple generalizations thereof). Of course, for most applications [1], [2] one needs to consider much more complicated models, however many important issues in string theory are still not understood, and the hope is that the two dimensional theory will serve as a useful toy model, in which some of these issues may be addressed. Due to the low dimension of space-time, the number of degrees of freedom in the theory is vastly smaller than in the twenty six (or ten) dimensional case. The physical on shell states include one field theoretic degree of freedom, the “tachyon” center of mass of the string (which is actually massless in two dimensions), and a discrete infinite set of massive states – the remnants of the tower of oscillator states in $D > 2$ (where the number of field theoretic degrees of freedom with mass $\leq m_0$ diverges exponentially with m_0).

Due to the absence of transverse excitations, one might expect the dynamics to be simpler in two dimensions. Indeed, following ideas of [3] [4] it has been shown [5] [6] using large N matrix model techniques, that many properties of two dimensional (2d) strings are exactly calculable to all orders in the topological (genus) expansion. The simplicity of these models is closely related to an underlying free fermion structure [3] – [7].

There are two main potential domains of application of the results of [5], [6]. One is two dimensional (world sheet) quantum gravity. The implications of matrix models for quantum gravity have been recently discussed in [8], [9], [10]. The second application, which will be of main interest to us here, is to critical (unified) string theory. The two are of course closely connected, and we will mention some aspects of this relation as we go along. The approach we will take is to try and understand the matrix model results in the continuum path integral formalism [1], and use this understanding to shed light on the structure of the theory, in particular its space-time dynamics. This program is far from complete; we will review what has been achieved and discuss some of the open problems.

We will take a broader point of view, presenting 2d strings in the context of higher dimensional string theory, which is ultimately what we’re interested in. This will also help to try and identify features of general validity, and those that are special to two dimensions. With that in mind, we will also investigate some string models in two dimensions which are more difficult to treat (or in some cases even formulate) using matrix models.

In the continuum formulation of 2d quantum gravity [11] [12], we are given a general conformal field theory (CFT) with action $\mathcal{S}_M(g)$ (on a Riemann surface with metric g_{ab}),

central charge c_M , and we have to integrate over all metrics g on manifolds of fixed topology (genus). In the conformal gauge $g_{ab} = e^\phi \hat{g}_{ab}$, the dynamical metric g is represented by the Liouville mode ϕ ; the action is:

$$S = S_L(\hat{g}) + S_M(\hat{g}) \quad (1.1)$$

where

$$S_L(\hat{g}) = \frac{1}{2\pi} \int \sqrt{\hat{g}} [\hat{g}^{ab} \partial_a \phi \partial_b \phi - \frac{Q}{4} \hat{R} \phi + 2\mu e^{\alpha_+ \phi}] \quad (1.2)$$

Q and α_+ are parameters which are fixed by gauge invariance [13]. It is very useful to think about the Liouville mode as a target space coordinate, so that (1.1) is a particular background of the critical string system in two dimensions [14]. Therefore, we have to set the total central charge of matter (c_M) and Liouville ($c_L = 1 + 3Q^2$) to 26:

$$Q = \sqrt{\frac{25 - c_M}{3}} \quad (1.3)$$

Spinless matter primary operators of dimension $\Delta (= \bar{\Delta})$ V_Δ acquire Liouville dependence,

$$T_\Delta(\phi) = V_\Delta \exp(\beta_\Delta \phi) \quad (1.4)$$

with

$$\Delta - \frac{1}{2}\beta(\beta + Q) = 1 \quad (1.5)$$

In particular, putting $\Delta = 0$ in (1.5) fixes α_+ in (1.2). As usual in critical string theory, (1.5) is the mass shell condition (or the linearized equation of motion for small excitations). The vertex operators T_Δ are related to the wave functions of the corresponding states Ψ_Δ by $T_\Delta(\phi) = g_{st}(\phi) \Psi_\Delta(\phi)$, where unlike the "usual" cases [1], the string coupling g_{st} is *not* constant in general (1.2); rather we have $g_{st} = g_0 \exp(-\frac{Q}{2}\phi)$. Therefore,

$$\Psi_\Delta = V_\Delta e^{(\beta + \frac{Q}{2})\phi} \quad (1.6)$$

and $E = \beta + \frac{Q}{2}$ is identified as the Liouville momentum of the state described by Ψ . The form (1.6) is actually only an asymptotic approximation of the exact wave function. It is valid in the region $e^{\alpha_+ \phi} \ll 1$ ($\phi \rightarrow \infty$ in our conventions). The exact wave function in two dimensional string theory, is known [9] to satisfy the Wheeler de Witt equation,

$$\left[-\frac{\partial^2}{\partial \phi^2} + \mu \exp(\alpha_+ \phi) + \nu^2 \right] \Psi_\Delta(\phi) = 0 \quad (1.7)$$

where $\nu^2 = 2\Delta - \frac{c_M - 1}{12}$. Eq. (1.7) was derived [9] in the matrix model approach. In the continuum formalism, (1.7) holds in a minisuperspace approximation; it was not yet derived in the full theory. It is very natural to interpret ϕ as Euclidean time. Eq. (1.5) takes then the form:

$$E^2 = m^2; \quad m^2 = 2\Delta - \frac{c_M - 1}{12} \quad (1.8)$$

The wave function Ψ_Δ with energy E describes one quantum mechanical degree of freedom in space-time. If, as in the minimal models of [15], the matter system contains a finite number of spinless primaries V_Δ , the corresponding string theory reduces to quantum mechanics of a finite number of degrees of freedom. In general, of course, the number of matter primaries is enormous, and the number of degrees of freedom is much larger than in field theory. If there are some non compact dimensions X_i , such that $V_\Delta(X) = e^{ik \cdot X} \tilde{V}_h$, $\Delta = \frac{1}{2}k^2 + h$, we can rewrite (1.8) as

$$E^2 = k^2 + m^2, \quad m^2 = 2h - \frac{c_M - 1}{12}$$

and V describes one field theoretic degree of freedom in the appropriate dimension.

From (1.8) we see that operators with $\Delta < \frac{c_M - 1}{24}$ in the matter CFT correspond in space-time to on shell tachyons, which lead (as in field theory) to IR instabilities. In field theory, these IR divergences imply that we are expanding around the wrong vacuum, and should move to a new, stable one. We will see that in string theory, as it is currently understood, the situation is surprisingly different.

From the point of view of 2d gravity, tachyons correspond to normalizable wave functions (1.6), while states with positive m^2 are described by wave functions which are exponentially supported at $\phi \rightarrow \pm\infty$ (and are of course non normalizable). The world sheet interpretation of this [8] is that normalizable (tachyonic) wave functions are supported mainly on surfaces with finite size holes in the *dynamical metric* g_{ab} ; they describe macroscopic states, and if we perturb the action (1.1) by the corresponding operators, the dynamical surface is destabilized by the multitude of holes that are created. Non normalizable wave functions with $E > 0$ are supported on very small surfaces in the dynamical metric, and therefore describe small disturbances of the surface (microscopic states). The corresponding operators can be (and were) studied in the matrix model approach. Wave functions with $E < 0$ describe very large disturbances of the surface, and it is not known how to treat them in the matrix model (although there are some plausible suggestions).

Thus, there is a nice correspondence between IR instabilities in space time (tachyons) and on the world sheet (macroscopic states which create holes in the surface). The role of states with $E < 0$ is not well understood at this time, and is in fact one of the important remaining problems, which is related to many properties of the theory, as we will see below. Note that by (1.8) the familiar relation between tachyons and relevant matter operators is only true for $c_M = 25$. In general, for $c_M < 25$ relevant operators with $\frac{c_M-1}{24} < \Delta < 1$ do not create instabilities; for $c_M > 25$ we have the unintuitive situation that irrelevant matter operators with $1 < \Delta < \frac{c_M-1}{24}$ are tachyonic. The region $c_M > 25$ has been less studied than $c_M < 25$, and we will not discuss it further here.

For tachyonic operators, β in (1.4) is complex (1.8). In the original work of [12], [13] (see also [16] [17]) it was noted that the identity operator ($V_\Delta = 1$ in (1.4)) becomes tachyonic when $c_M > 1$. Since the gravitational coupling constant is proportional to $\frac{1}{26-c_M}$ [11], it has been suggested that at $c_M = 1$ a certain transition between “weak gravity” $c_M < 1$ and “strong gravity” $c_M > 1$ takes place. Taken naively, this point of view was not completely satisfactory, since it is easy to construct matter theories where $c_M < 1$, but there exist tachyonic operators in the spectrum; it is not likely that the identity operator plays a special role in this context. The issue was later clarified in [8], [18]; the current understanding is that the crucial quantity is not the gravitational coupling constant, but rather the *density of states* of the theory. Gravity has $\frac{1}{2}n(n-3)$ degrees of freedom in n dimensions; for $n = 2$ this is -1 . If the matter theory has more than “one degree of freedom”, in a sense which we will make precise below, the full string theory contains tachyons. Therefore, the invariant way to describe the famous “ $c = 1$ barrier” of [12], is that *2d gravity coupled to matter with more than one field theoretic degree of freedom is unstable*. The relation between the existence of tachyonic excitations around a certain vacuum, and the number of states in it implies that if it is possible at all to turn on expectation values of the various fields and move from a tachyonic vacuum to a stable one in string theory, the new stable vacuum is always “trivial” (at least in bosonic string theory).

Hence, the density of states is of major importance in string theory, as are tachyonic instabilities. Section 2 of these notes is devoted to a precise definition of the density of states and its relation to the presence of tachyons and stability. We also discuss the implications of this relation for string dynamics.

Since the only stable vacua of bosonic string theory (where by bosonic we mean vacua with only bosonic space-time excitations; e.g. the fermionic string falls into this category

as well) are those which are essentially two dimensional field theories in space-time, we are led in section 3 to consider theories with space-time fermions [19]. We construct explicitly a large set of theories with a number of degrees of freedom varying between that of a two dimensional field theory and that of the ten dimensional superstring, all of which are stable solutions of the equations of motion of string theory. This is possible since space time fermions contribute a negative amount to the “number of degrees of freedom”, and a theory with many bosons and many fermions can still have a vanishingly small total “density of states”. This number is required to be small for stability. The theories thus obtained have many of the favorable properties of critical superstring models, but exhibit some puzzling features as well (such as continuous breaking of SUSY), mostly related to their time dependence.

In section 4 we describe the known facts about classical dynamics in two dimensional string theory. One of the issues of interest is the exact form of the classical equations of motion of the string. The usual way one obtains those in string theory is by studying scattering amplitudes of on shell fields. Calculating the correlation functions of T_{Δ} (1.4) involves solving the interacting Liouville theory (1.2). Despite many efforts [16], [17], [8], [20], this is still an open problem. What saves the day is that there exists a class of amplitudes which contain most of the information, and are simply calculable. To understand this, it is useful to focus on the dynamics of the zero mode of the Liouville field ϕ (the minisuperspace approximation in 2d gravity). The main source of the complications in Liouville theory is the “cosmological” interaction in (1.2). From the space-time point of view, this corresponds to a (zero X momentum) tachyon condensate, while on the world sheet this is a potential for ϕ . The reason why it is needed is to keep the Liouville field away from the region $\phi \rightarrow -\infty$ where the string coupling $g_{st} \rightarrow \infty$. Many amplitudes diverge when $\mu \rightarrow 0$ in (1.2). By KPZ scaling [12], the amplitudes $\langle T_{\Delta_1} \dots T_{\Delta_n} \rangle$ are proportional to μ^s , where

$$\sum_{i=1}^N \beta_i + \alpha_+ s = -Q \quad (1.9)$$

In general, all amplitudes are non zero – there is no Liouville momentum conservation (1.9). Instead, the ϕ zero mode integral, which can be explicitly performed, yields [21]:

$$\langle T_{\Delta_1} \dots T_{\Delta_n} \rangle = \left(\frac{\mu}{\pi} \right)^s \Gamma(-s) \langle T_{\Delta_1} \dots T_{\Delta_n} \left[\int \exp(\alpha_+ \phi) \right]^s \rangle_{\mu=0} \quad (1.10)$$

where the correlator on the r.h.s. is understood to exclude the Liouville zero mode. One can see from (1.10) that the amplitudes with $s = 0$ are special. The ϕ zero mode integral is given for them by:

$$\int_{-\infty}^{\infty} d\phi_0 \exp(-\mu e^{\alpha+\phi_0})$$

This diverges from $\phi \rightarrow \infty$; imposing a UV cutoff $\phi_0 \leq \phi_{UV}$, we find $\log \mu$ multiplied by a free field amplitude for the Liouville field (i.e. Liouville is treated as a Feigin-Fuchs field, and the cosmological interaction is absent). Of course, we still have to multiply by the matter contribution and integrate over moduli, but this reduces the calculation to well defined integrals; similar representations for generic Liouville amplitudes (1.10) are not known.

Why does this simplification of the amplitudes with $s = 0$ occur? A major clue is provided by the observation that these amplitudes are insensitive to the particular form of the “wall” that keeps the Liouville field away from $\phi = -\infty$. For example, replacing $e^{\alpha+\phi}$ by a general T_Δ would do the job, and while generic amplitudes depend strongly on the choice of the condensate T_Δ , the $s = 0$ ones do not (apart from trivial overall factors). The existence of the Liouville wall is of course a boundary effect, hence the $s = 0$ amplitudes correspond to *bulk scattering*, which is clearly independent of the form of the boundary at strong coupling. The reason for the simplicity is that in the bulk of the ϕ volume, all the complications of Liouville theory are irrelevant, and the Liouville field is free. The overall factor of $\log \frac{1}{\mu}$ which appears in all bulk amplitudes is interpreted as the volume of the Liouville coordinate, as appropriate for bulk effects.

A similar simplification occurs in (1.10) for all $s \in Z_+$. The power of $\int \exp(\alpha_+ \phi)$ on the r.h.s. becomes then a positive integer; the divergence of $\mu^s \Gamma(-s)$ is interpreted again as¹: $\mu^s \Gamma(-s) \simeq \mu^s \log \mu$. The space-time interpretation is as before: now it involves bulk interaction of the scattering particles $T_{\Delta_1} \dots T_{\Delta_n}$, with the zero momentum tachyons that form the Liouville wall.

The bulk amplitudes provide us with the required probe of the dynamics of D dimensional string theories; they can be defined and evaluated for all string vacua (all matter

¹ This can be obtained by starting with the fixed area amplitude, using $\mu^s \Gamma(-s) = \int_0^\infty dA A^{-s-1} e^{-\mu A}$. The divergence at $s \in Z_+$ is interpreted then as a small area divergence in the integral over areas, which can be regulated by a small area (UV) cutoff, as before, yielding the stated result.

CFT's M), while general Liouville amplitudes (1.10) are expected to be much more subtle for $D > 2$. The bulk amplitudes should also be sufficient to deduce the equations of motion of the theory, and compare the dynamics around different vacua. In principle one can construct from them a space-time Lagrangian which gives those amplitudes, and then incorporate boundary effects to deduce all the amplitudes of the theory [22].

In view of the above discussion, section 4 is devoted to the bulk tree level S - matrix in two dimensional string theory. Surprisingly, the moduli integrals can be explicitly evaluated in this case, essentially due to the same simplification of the dynamics of the theory, as that leading to the free fermion structure of the corresponding matrix models. In the continuum formalism, the simplicity is due to a partial decoupling of a certain infinite set of discrete string states. We review the results and mention some problems still left in the continuum approach to 2d strings. We also compare the situation in 2d closed string theory to that found in open 2d string theory. Section 5 contains some concluding remarks.

2. DENSITY OF STATES AND TACHYONS IN STRING THEORY.

In this section we will show that the density of states (or number of degrees of freedom) of the system plays a central role in two dimensional quantum gravity and string theory. In particular, we will show that matter with 'too many' degrees of freedom can not be consistently coupled to gravity (under mild assumptions) due to the appearance of tachyons. We believe that this restriction on the number of states in quantum mechanical generally covariant theories is more general.

We start by defining the density of states. The most general situation we are interested in is an arbitrary vacuum of critical string theory, i.e. an arbitrary conformal field theory with $c = 26$ (or $\hat{c} = 10$). As we saw in section 1, gravity coupled to a matter CFT with any c_M is a particular example, with the missing central charge $26 - c_M$ supplied by the Liouville CFT. We'll start by considering this special case, and then generalize. To count states it is convenient to evaluate the torus partition sum:

$$Z(\tau) = \text{Tr} q^{L_0 - \frac{c}{24}} \bar{q}^{\bar{L}_0 - \frac{\bar{c}}{24}} \quad (2.1)$$

where $q = e^{2\pi i \tau}$, and $\tau = \tau_1 + i\tau_2$ is the complex modulus of the torus.

The string partition sum (2.1) factorizes into a product of three contributions, from ghosts (for which a factor of $(-)^G$, G - ghost number, must be inserted in (2.1)), Liouville and the "matter" CFT. The ghosts contribute to $Z(\tau)$ a factor of $|\eta(\tau)|^4$. The Liouville

contribution is that of a free scalar field [21] $\log \mu (\sqrt{\tau_2} |\eta(\tau)|^2)^{-1}$; the cosmological term in (1.2) enters trivially, for the same reasons as in section 1 – the one loop free energy is a bulk amplitude; hence it is proportional to the volume $\log \mu$, with the coefficient given by a free field amplitude. Finally, the matter is described by a partition sum $Z_{\text{CFT}}(\tau)$ (defined as in (2.1)).

The total partition sum is (after multiplying by a factor of τ_2 for later convenience)

$$Z_{\text{string}} = \sqrt{\tau_2} |\eta|^2 Z_{\text{CFT}} \quad (2.2)$$

Modular invariance of $Z_{\text{string}}(\tau)$ is a fundamental principle in string theory. We will assume it throughout our discussion. In ordinary CFT, the partition sum (2.1) can be used to count states in the theory. One considers the behavior of the partition sum (Z_{CFT} or Z_{string}), when $\tau_1 = 0, \tau_2 = \beta \rightarrow 0$, where $Z(\beta) = \text{Tr } e^{-\beta E}$ has the form:

$$Z(\beta) \sim \beta^r \exp\left(\frac{\pi c_{\text{eff}}}{6\beta}\right) \quad (2.3)$$

Clearly, c_{eff} is a measure of the number of degrees of freedom of the theory, since in the limit $\beta \rightarrow 0$ all states contribute 1 to $Z(\beta)$ (2.1). By modular invariance $Z(\beta) = Z(\frac{1}{\beta})$, and (2.1), c_{eff} is related to the lowest lying states in the spectrum:

$$\begin{aligned} c_{\text{eff}} &= c - 24\Delta_m \\ \Delta_m &= \frac{1}{2} \min(\Delta + \bar{\Delta}) \end{aligned} \quad (2.4)$$

In unitary CFT $\Delta_m = 0$, and $c_{\text{eff}} = c$. So the central charge measures the number of degrees of freedom [23]. In non unitary theories, Δ_m is generically negative and $c_{\text{eff}} > c$. In Liouville theory, $c_{\text{eff}} = 1$ ($\Delta_m > 0$). Thus $c (= 1 + 3Q^2)$ is not a good measure of the density of states, which is that of a free scalar field for all Q .

In string theory, c_{eff} is not a good measure of the density of states since not all states contributing to (2.3) are physical – we have to impose the constraint $L_0 = \bar{L}_0$. This is implemented by integrating over τ_1 ; we define:

$$G(\tau_2) = \int_{-\frac{1}{2}}^{\frac{1}{2}} d\tau_1 Z_{\text{string}} \quad (2.5)$$

By modular invariance we know that $Z_{\text{string}}(\tau_1 + 1, \tau_2) = Z_{\text{string}}(\tau_1, \tau_2)$, hence only states of integer spin contribute to Z_{string} ($\Delta - \bar{\Delta} \in \mathbb{Z}$). Transforming to $G(\tau_2)$ (2.5) eliminates

contributions of all states except those with $\Delta = \bar{\Delta}$. The function $G(\tau_2)$ has a simple space-time interpretation. It is related to the one loop free energy of the particle excitations of the string: $\text{Tr}(-)^F \log(p^2 + m^2)$ (F is the space-time fermion number). In a proper time representation, the one loop free energy Ω is given by:

$$\Omega = \text{Tr}(-)^F \log(p^2 + m^2) = \int_0^\infty \frac{ds}{s} \sum_n (-1)^{F_n} \int d^D p e^{-s(p^2 + m_n^2)} \quad (2.6)$$

where p is the momentum in the non compact directions and the sum over n runs over the space-time excitations. The function $G(\tau_2)$ of (2.5) is related to the integrand in (2.6):

$$G\left(\frac{s}{2\pi}\right) = s \sum_n (-1)^{F_n} \int d^D p e^{-s(p^2 + m_n^2)} \quad (2.7)$$

One consequence of this is that all states should contribute positive amounts to $G(\tau_2)$, unless there are space-time fermions in the spectrum. From the definition of $G(\tau_2)$ (2.2), (2.5), this is far from clear. While the CFT partition sum as well as that of Liouville satisfy this property, being traces over Hilbert spaces with positive weights, ghost oscillators flip the sign of the contributions to (2.1) (due to the factor of $(-)^G$ in the definition of the ghost partition sum). Therefore, even in bosonic string theory there may appear physical states at non trivial ghost numbers, which contribute a negative amount to $G(\tau_2)$ (2.5). Such states indeed do appear at discrete values of the momenta p . They seem to decouple from the dynamics due to the ghost numbers, and their role is not completely clear. In more than two dimensions the statistical weight of these states is low and they can be ignored (except perhaps as generators of symmetries). In two dimensional space-time they seem to be closely related to the symmetry structure of the theory [24]. In any event, the field theoretic degrees of freedom, over which the sum in (2.7) runs, occur at zero ghost number and contribute with positive sign to the partition sum.

A state with $\Delta = \bar{\Delta}$ in the CFT contributes to $G(\tau_2)$, an amount $e^{-4\pi\tau_2(\Delta - \frac{cM-1}{24})}$. If $\Delta < \frac{cM-1}{24}$, the Schwinger integral (2.6) develops an IR divergence (from $\frac{s}{2\pi} = \tau_2 \rightarrow \infty$). This IR divergence is due to the tachyon instability. Hence, the behavior of $G(\tau_2)$ as $\tau_2 \rightarrow \infty$ probes the existence of tachyons in the spectrum. On the other hand, its behavior as $\tau_2 \rightarrow 0$ is a measure of the density of states of the theory, since in that limit all states contribute 1 to G . Now, the theories we are describing here all have an infinite number of states, therefore generically G diverges as $\tau_2 \rightarrow 0$; however we can estimate the density of states by measuring how fast G diverges in that limit.

On general grounds we know that as $\tau_2 \rightarrow 0$, $G(\tau_2) \simeq \tau_2^2 e^{\frac{\pi}{\tau_2}}$. Therefore it is natural to define:

$$c_{\text{string}} - 1 = \lim_{\tau_2 \rightarrow 0} \frac{6\tau_2}{\pi} \log G(\tau_2) \quad (2.8)$$

as a measure of the number of degrees of freedom. The -1 on the left hand side of (2.8) reflects the contribution of the gravity sector to the density of states mentioned above. If one takes, e.g., the matter CFT to be one scalar field or a minimal model [15], it is easy to see that $c_{\text{string}} = 1$. Thus the minimal models and two dimensional string theory have the same density of states. In general, $c_{\text{string}} > 1$ for theories without space-time fermions. We will see soon that such theories always contain tachyons.

As explained above, tachyons cause an IR divergence in Ω , thus finiteness of Ω is a good measure of the existence of tachyons (in fermionic string theories this is strictly true only for theories without tachyonic space-time fermions, e.g. theories with unitary matter). Of course, the field theoretic expression for Ω (2.6) has in addition to the IR tachyon divergence which we are interested in, a UV divergence (from $s = 0$). This is usually dealt with by a proper time cutoff: $\int_0^\infty ds \rightarrow \int_{\Lambda^{-2}}^\infty ds$. String theory removes this divergence by a different mechanism [1]. One notes that by (2.5), (2.6), (2.7), $\Omega = \int \frac{d^2\tau}{\tau_2^2} Z_{\text{string}}(\tau)$, where the integral runs over the half infinite strip $\tau_2 \geq 0$, $|\tau_1| \leq \frac{1}{2}$. This may diverge due to modular invariance of Z_{string} since the strip contains an infinite number of copies of a fundamental domain of the modular group. If this is the only source of divergence, one can cure it by restricting the integral to a fundamental domain $\mathcal{F}$ of the modular group, $|\tau| \geq 1$, thus avoiding the “UV region” $\tau_2 \rightarrow 0$. But this “stringy regularization” actually teaches us something very interesting about the theory. Suppose there are no tachyons in the spectrum. Then the integral $\int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} Z_{\text{string}}$ is finite, and the only possible source of divergence of the integral over the strip is the volume of the modular group. We can try to estimate this divergence by using the “field theoretic” cutoff: integrate $\int \frac{d^2\tau}{\tau_2^2}$ over the cutoff strip: $|\tau_1| \leq \frac{1}{2}$, $\tau_2 > \frac{1}{\Lambda^2}$. The integral of Z_{string} over the cutoff strip should diverge at the same rate as the (regularized) volume of the modular group, which is given by $\int_{\Lambda^{-2}}^\infty \frac{d\tau_2}{\tau_2^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} d\tau_1 1 = \Lambda^2$. Hence we learn that from the space-time point of view, tachyon free string theories have a free energy $\Omega(\Lambda)$, which diverges (at most) as $\Omega \simeq \Lambda^2$ as we remove the cutoff. But in space-time the rate of divergence of Ω measures the density of states of the theory. The behavior we find in tachyon free string theory indicates a very small number of states. Even a theory with one bosonic field in D dimensions would have $\Omega(\Lambda) \simeq \Lambda^D$. In string theory we generically have an infinite number of space-time fields,

so the behavior we find is even more unusual. We would expect $\Lambda^a e^{b\Lambda}$ in general, due to the Hagedorn-like growth of the number of states with mass. By the above arguments, such theories are always tachyonic. The only theories that can be tachyon free are those that exhibit the general features of 2D field theories!

Pushing these ideas one step further, we can derive a more quantitative correspondence between Ω and the number of states. If the divergence of (2.6) is indeed due to the volume of the modular group, we should have a relation of the form:

$$\lim_{\Lambda \rightarrow \infty} \frac{\int_{\Lambda^{-2}}^{\infty} \frac{d\tau_2}{\tau_2^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} d\tau_1 Z_{\text{string}}}{\int_{\Lambda^{-2}}^{\infty} \frac{d\tau_2}{\tau_2^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} d\tau_1 1} = \frac{\int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} Z_{\text{string}}}{\int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} 1} \quad (2.9)$$

Equation (2.9) is the statement that the two regularizations of Ω using the field theoretic proper time cutoff, and the stringy modular invariant cutoff, are equivalent. This is in fact only true under certain assumptions, which we will soon state, but assuming it is true, we derive by evaluating the l.h.s. the following relation:

$$\int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} Z_{\text{string}} = \lim_{\Lambda \rightarrow \infty} \frac{\pi}{3\Lambda^2} \int_{\Lambda^{-2}}^{\infty} \frac{d\tau_2}{\tau_2^2} G(\tau_2) = \lim_{\Lambda \rightarrow \infty} \frac{\pi}{3} G(\Lambda^{-2}) \quad (2.10)$$

This relation states that the (regularized) one loop free energy is equal (up to a constant) to the number of states of the theory $G(0)$. Before going on to prove (2.10), we would like to make several comments on its significance and implications.

Eq. (2.10) is a statement about modular invariant functions (obeying certain conditions to be stated below). We have presented it for the case of a CFT coupled to Liouville, however the discussion applies to all string vacua which are described by modular invariant partition sums. Important generalizations include fermionic strings, superstrings, and heterotic strings, as well as arbitrary CFT's with $c = 26$.

The l.h.s. of (2.10) diverges iff there are tachyons in the spectrum. Hence, if there are no tachyons, the r.h.s. must also be finite. Looking back at (2.7), we see that this implies that the theory has the number of states of a two dimensional field theory with a finite number of fields. If there are no space-time fermions, this means that the theory contains far fewer states than generic string theories [1]. The only bosonic string theories without tachyons are two dimensional (e.g. the $c = 1$ or minimal models coupled to gravity, and the coset model of [25]). This means that the role of tachyons in string theory is more fundamental than in field theory. Unlike there, it doesn't seem likely that tachyons can be gotten rid of by shifting to a nearby vacuum. For that to happen, one of two scenarios

must occur: either space-time fermions must somehow dynamically appear, or the new stable vacuum must effectively be a two dimensional string theory. Starting e.g. from the 26 dimensional bosonic string, neither possibility seems likely. In general, it seems unlikely that turning on expectation values of fields can change c_{string} , which should be invariant under small perturbations. The fate of string vacua containing tachyons (or 2D quantum gravity systems with “too much” matter) remains unclear.

Fermionic string theories with a non chiral GSO projection contain also only bosonic excitations in space-time. Hence the situation there is completely analogous to the bosonic case – tachyon free theories have very few states (the number of states is again as in two dimensional field theory). Theories with many states contain tachyons. It is well known that non trivial theories exist in this case, but those involve a *chiral* GSO projection (e.g. the ten dimensional superstring [1]). It is important to note that such theories are not necessarily space-time supersymmetric (one example of a non supersymmetric theory is the $O(16) \times O(16)$ theory [1]). However, it is known that they always contain space-time fermions. Our discussion uncovers a surprising feature of these theories: even though space-time SUSY is absent in general, tachyon free superstrings are *asymptotically supersymmetric*. In other words, while the number of bosons and fermions is not the same energy level by energy level, at high enough energy, the total numbers of bosons and fermions up to that energy are the same to fantastic precision. More precisely, in this case it is natural to write $G(s)$ (2.7) as $G(s) = G_B(s) - G_F(s)$, where G_B and G_F measure the contributions of bosons and fermions to the free energy (2.7), and the relative minus sign is due to spin statistics; then in general we have $G_B(s), G_F(s) \simeq s^a e^{\frac{1}{s}}$, while $G_B(s) - G_F(s) \rightarrow \text{const}$ as $s \rightarrow 0$ (again, by using (2.10) and finiteness of its l.h.s.). Hence bosons and fermions almost cancel in any tachyon free string theory. This “asymptotic supersymmetry” is a very puzzling phenomenon, whose full implications are still not understood. Non supersymmetric tachyon free superstring theories resemble models with spontaneously broken SUSY (of course the scale of breaking is the Planck scale in general). Whether this is more than a formal similarity remains to be seen. Note also that the one loop cosmological constant in tachyon free string theories is unnaturally small from the space-time point of view (although still much too large). From our point of view, Ω can vanish without space-time supersymmetry. This requires only that G_B and G_F cancel precisely as $s \rightarrow 0$. We don’t know whether or why this should be the case.

In theories with space-time fermions, one may also define $\tilde{G}(s) = G_B(s) + G_F(s)$, which counts the total density of states of bosons *plus* fermions. This quantity behaves

when $s \rightarrow 0$ as: $\tilde{G}(s) \simeq s^a e^{\frac{1}{2}}$, and determines the thermal properties (e.g. the Hagedorn temperature) of the theory. It is not clear whether models with $b > 0$ (such as those of section 3) exhibit any simplifying features due to the fact that $G_B(s) - G_F(s) \simeq 0$ (as $s \rightarrow 0$).

Our remaining task in this section is to state precisely and prove (2.10).

Theorem: Let $Z_{\text{string}}(\tau)$ be a modular invariant function, which is finite throughout the fundamental domain $\mathcal{F}$, except perhaps at $\tau_2 = \infty$, such that:

- 1) $\int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} Z_{\text{string}} = \text{finite}$. It is implied in the definition of the above integral that for $\tau_2 \gg 1$ we first perform the integral over $-\frac{1}{2} \leq \tau_1 \leq \frac{1}{2}$, and then integrate over τ_2 .
- 2) As $\tau_2 \rightarrow \infty$, $Z_{\text{string}} \simeq \tau_2^a q^a \bar{q}^b$ ($q = e^{2\pi i \tau}$) where $a, b > -1$.

Then:

$$\int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} Z_{\text{string}} = \frac{\pi}{3} \lim_{\tau_2 \rightarrow 0} \int_{-\frac{1}{2}}^{\frac{1}{2}} d\tau_1 Z_{\text{string}}(\tau_1, \tau_2) \quad (2.11)$$

Proof: Consider the function:

$$F(R) = \int_0^\infty \frac{d\tau_2}{\tau_2^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} d\tau_1 Z_{\text{string}}(\tau) \sum_{r \neq 0} \exp\left(\frac{-\pi R^2 r^2}{\tau_2}\right) \quad (2.12)$$

This is a cutoff version of (2.6), with a “modular invariant” cutoff. R plays the role of Λ^{-1} in the discussion above. Our purpose is to relate the integral over the strip $F(R)$ (2.12) to an integral over the fundamental domain $\mathcal{F}$. Naively, this can be done using [26]:

$$\int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} Z_{\text{string}}(\tau) \exp\left(-\pi R^2 \frac{|n - m\tau|^2}{\tau_2}\right) = \int_{\alpha \cdot \mathcal{F}} \frac{d^2\tau}{\tau_2^2} Z_{\text{string}}(\tau) \exp\left(-\pi R^2 \frac{\tau^2}{\tau_2}\right) \quad (2.13)$$

where α is the modular transformation taking $e^{-\pi R^2 \frac{|n - m\tau|^2}{\tau_2}} \rightarrow e^{-\pi R^2 \frac{\tau^2}{\tau_2}}$. We can write the sum over r and integral over the strip (2.12) as a sum over r, α and an integral over the fundamental domain $\mathcal{F}$ (since the sum over $\alpha \cdot \mathcal{F}$ generates the strip), and then replace the sum over r, α by a sum over n, m (2.13) (and integral over $\mathcal{F}$). This would suggest:

$$\int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} Z_{\text{string}}(\tau) \sum_{n, m \in \mathbb{Z}} \exp\left(-\pi R^2 \frac{|n - m\tau|^2}{\tau_2}\right) = F(R) + \int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} Z_{\text{string}} \quad (2.14)$$

This is unfortunately too naive. The problem is that in many interesting situations (most notably heterotic strings), the theory contains “unphysical tachyons” (with $L_0 \neq \bar{L}_0$), which lead to divergent terms as $q \rightarrow 0$, but disappear after the τ_1 integration. In such situations the above integrals (e.g. (2.13)) are not absolutely convergent. We have to

integrate the l.h.s. of (2.13) over τ_1 first. In such cases, the simple order of integration on $\mathcal{F}$ translates to a complicated prescription on $\alpha \cdot \mathcal{F}$, different for each modular transformation α . Since the integral depends on the order of integration, (2.14) is in general invalid. Clearly, in general the situation is out of control and the argument of [26] can only be applied if we're dealing with absolutely convergent integrals. Fortunately, this is the case if R in (2.14) is large enough. By condition (2) of the theorem, we are dealing with $Z_{\text{string}}(\tau)$ such that there exists a R_0 such that for all $R > R_0$ the l.h.s. of (2.13) is independent of the order of integration, as long as $m \neq 0$. But, since we use (2.13) only for terms with $m \neq 0$ (otherwise $\alpha = 1$ in (2.13)), for $R > R_0$, (2.14) is valid. Now we can continue (2.14) analytically to $R < R_0$ and it must still hold. The reason is that both the l.h.s. and the r.h.s. of (2.14) define analytic functions of R in a strip $\text{Im}R < \epsilon$, $R > 0$. For the l.h.s. this is clear, since the only divergences come from $\tau_2 \rightarrow \infty$, but by condition (2) of the theorem no such divergence is possible for the function and all its derivatives w.r.t. R . The r.h.s. is analytic since it is equal to the l.h.s. for $R > R_0$, and the analytic continuation is unique. We conclude that (2.14) is correct all the way down to $R = 0$. Poisson resummation of the l.h.s. gives

$$\frac{1}{R} \int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^3} Z_{\text{string}}(\tau) \sum_{n,m \in \mathbb{Z}} q^{\frac{1}{4}(\frac{n}{R} + mR)^2} \bar{q}^{\frac{1}{4}(\frac{n}{R} - mR)^2} = F(R) + \int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^2} Z_{\text{string}}(\tau) \quad (2.15)$$

As $R \rightarrow 0$ both sides behave as $\frac{1}{R^2}$. Equating the coefficients gives (2.11).

The conditions of the theorem are very natural. Condition 1 is obvious; the main reason to require condition 2 is that it is equivalent to unitarity when $c < c_{\text{critical}}$. E.g. in (2.2) we have $Z_{\text{string}} \simeq q^{\Delta - \frac{c_M - 1}{24}} \bar{q}^{\tilde{\Delta} - \frac{c_M - 1}{24}}$ (as $q \rightarrow 0$), and as long as $c_M < 25$, if all $\Delta \geq 0$ (as is the case in unitary CFT's), condition 2 is satisfied. The reason one expects the theorem to break down in general for non unitary matter, is that then the Ramond sector may contain tachyons as well as the NS sector. But in that case, bosonic and fermionic tachyons may cancel in the expression for the free energy, giving rise to a finite l.h.s. of (2.11), but still contribute to a large density of states, such that the r.h.s. of (2.11) is infinite. For $c > c_{\text{critical}}$, even unitary theories may violate the theorem. This region is not understood, as mentioned above. Some subtleties may also occur on the boundary between the two regions – the critical (heterotic) string (see [18] for details).

3. NON – CRITICAL SUPERSTRINGS.

In section 2 we saw that all non trivial theories of strings without tachyon instabilities must contain space-time fermions, in fact essentially the same number of fermions and of

bosons. In the critical (“ten dimensional”) case, this is achieved by a chiral GSO projection [1]. In this section we will see that a straightforward generalization exists in the non-critical case as well, giving rise to a large class of stable string theories with the number of degrees of freedom varying between that of two dimensional string theory and that of the critical superstring. We will discuss the “type II” theory – as usual there exists a heterotic version.

Let us first recall how the GSO projection works in the case of ten dimensional (perhaps compactified) string theory. One starts with a fermionic string vacuum, which is normally left-right symmetric (non chiral GSO projection), and therefore contains two sectors (NS, NS) and (R, R) (for the left and right movers). Both sectors contain bosonic excitations in space-time – all states contribute with positive sign to the partition sum. As discussed above, such theories are tachyonic due to the large number of states ($c_{\text{string}} > 1$). A chiral GSO projection can be implemented if the theory has a chiral Z_2 R – symmetry. If there are no global anomalies, we can gauge this Z_2 symmetry (or in other words orbifold by it [1]); this involves two elements; first eliminating all states which are not invariant under the Z_2 symmetry from the spectrum. Hence only a subset of the original bosonic (NS, NS) and (R, R) states survive the projection. Then, as standard for orbifolds, there are also twisted sectors, which in this context are the (NS, R) and (R, NS) sectors, which contain space-time fermions. The theory thus obtained is not necessarily space-time supersymmetric, however under favorable circumstances it is tachyon free.

It is not known in general which Z_2 symmetries can be gauged and furthermore give rise to tachyon free “superstring” models. There is, however, a large class of theories, where this Z_2 is part of a larger chiral algebra, which are in general tachyon free and possess some additional nice properties. These are the space-time supersymmetric string vacua. In general such vacua can be constructed iff the original fermionic string vacuum has a global $N = 2$ superconformal symmetry². The projection can then be performed by constructing the (chiral) space-time SUSY charge S [28], and projecting out all operators which are not local w.r.t. S ; then one should again add “twisted sectors” obtained by acting with S on the remaining (NS, NS) and (R, R) operators; this gives rise to the (R, NS) and (NS, R) sectors (space-time fermions). This procedure is expected to be in general

² To avoid confusion it is important to emphasize that this $N = 2$ symmetry is an “accidental” global symmetry of the vacuum. It is not gauged (the BRST charge is that of the $N = 1$ fermionic string), and most excitations are not invariant under this symmetry – it is a property of the vacuum and not of the full theory. The existence of the $N = 2$ symmetry is a necessary and sufficient condition for space-time SUSY [27].

free of global anomalies (modular invariant), although no general proof exists. Also, such vacua are automatically free of tachyons, assuming there is only one time direction and the remaining “matter” is unitary. This follows from the space-time SUSY algebra. We will not elaborate on this procedure here, but rather explain directly how it generalizes to the non-critical case. For details on the “critical” construction we refer the reader to the original literature [28], [27], [29].

The procedure we are going to describe turns (non-critical) fermionic string vacua with an accidental global $N = 2$ symmetry to consistent superstrings. We will describe it in the particular case of the $D = d + 1$ dimensional fermionic string. The general structure is an obvious generalization, and can be found in [19].

The $d + 1$ dimensional fermionic string is described in superconformal gauge by d matter superfields X_i , $i = 1, \dots, d$, and the super Liouville field Φ . In components we have:

$$\begin{aligned} X_i &= x_i + \theta\psi_i + \bar{\theta}\bar{\psi}_i + \theta\bar{\theta}F_i; \quad i = 1, \dots, d \\ \Phi &= \phi_l + \theta\psi_l + \bar{\theta}\bar{\psi}_l + \theta\bar{\theta}F_l \end{aligned} \quad (3.1)$$

The fields X_i are free, while Φ is described by the super Liouville Lagrangian [11]:

$$S_{SL} = \frac{1}{2\pi} \int d^2z \int d^2\theta [D\Phi\bar{D}\Phi + 2\mu \exp(\alpha_+\Phi)] \quad (3.2)$$

where $D = \partial_\theta + \theta\partial_z$, and we have dropped curvature couplings (1.2). The central charge is $\hat{c}_L (= \frac{2}{3}c_L) = 1 + 2Q^2$, $Q = \sqrt{\frac{9-d}{2}}$. As mentioned above, this theory has a tachyonic ground state (for $d > 1$). The corresponding vertex operator is

$$T_k = \int d^2\theta \exp(ik \cdot X + \beta\Phi) \quad (3.3)$$

where $\frac{1}{2}k^2 - \frac{1}{2}\beta(\beta + Q) = \frac{1}{2}$. As in the bosonic case, $E = \beta + \frac{Q}{2}$ (see section 1 (1.6)), $E^2 - k^2 = m^2 = \frac{1-d}{8}$, so that T_k is a tachyon for $d > 1$.

For $d = 9$ (critical ten dimensional string theory), the Z_2 R - symmetry that one orbifolds by to obtain the superstring is $\psi \rightarrow -\psi$, $\bar{\psi} \rightarrow \bar{\psi}$ (or equivalently $\theta \rightarrow -\theta$, $\bar{\theta} \rightarrow \bar{\theta}$). The tachyon (3.3) transforms to minus itself under this symmetry, and is projected out of the spectrum. It is tempting to try and divide by the same symmetry for all d . Unfortunately, this is too naive; for $d < 9$ the orbifoldized theory doesn't make sense; there are a number of ways of seeing that - modular invariance breaks down, the spin field [28] doesn't have the necessary properties, etc. We have to use the more sophisticated approach related to global $N = 2$ symmetry. The basic observation is that for odd $d = 2n + 1$, such

that the total dimension of space-time is even³, the system (3.1) possesses an $N = 2$ symmetry. This symmetry pairs $d - 1 = 2n$ of the X_i into $N = 2$ superfields, and the remaining $X_{2n+1} \equiv X$ is paired with Φ . This involves breaking the $O(d)$ symmetry; one can think of Φ, X as light cone coordinates, while the rest of the X_i ($i = 1, \dots, 2n$) are transverse directions. The vacuum is not translationally invariant in the Φ direction. The action of the $N = 2$ superconformal symmetry on the transverse X_i is completely standard and we will not review it. The free Φ, X system forms a chiral $N = 2$ multiplet: it is convenient to define $\phi = \phi_l + ix$, $\psi = \psi_L + i\psi_x$, etc. The $N = 2$ generators are:

$$\begin{aligned}
 T(z) &= -\frac{1}{2}\partial\phi\partial\phi^\dagger + \frac{1}{4}(\psi\partial\psi^\dagger + \psi^\dagger\partial\psi) - \frac{1}{4\gamma}\partial^2\phi - \frac{1}{4\gamma}\partial^2\phi^\dagger \\
 J(z) &= \frac{1}{2}\psi\psi^\dagger - \frac{1}{2\gamma}\partial\phi + \frac{1}{2\gamma}\partial\phi^\dagger \\
 T_F^+ &= \frac{1}{2}\psi^\dagger\partial\phi + \frac{1}{2\gamma}\partial\psi^\dagger \\
 T_F^- &= \frac{1}{2}\psi\partial\phi^\dagger + \frac{1}{2\gamma}\partial\psi
 \end{aligned} \tag{3.4}$$

In (3.4), $\dagger$ does not interchange left and right movers – it corresponds to complex conjugation in field space: $\phi^\dagger = \phi_l - ix$, etc. As in Liouville theory, the free field expressions (3.4) should be understood literally away from an interacting region (“wall”), which we haven’t specified yet. The string coupling is again (as in the bosonic theories) $g_{\text{st}} \propto e^{-\frac{2}{3}\phi}$, and we need a potential to suppress the region $\phi \rightarrow -\infty$; we haven’t yet determined the appropriate potential. The form (3.4) is still very useful, for example to discuss bulk effects, which are anyway the best understood part of the Liouville correlation functions.

To choose an appropriate wall, it is convenient to define the chiral $N = 2$ superfield $\Gamma = \phi + \theta\psi + \bar{\theta}\bar{\psi} + \theta\bar{\theta}F + \dots$ (such that $D^\dagger\Gamma = 0$). Here θ is a complex Grassmanian variable (different from the θ in (3.1) – (3.3)). In terms of Γ , the free Lagrangian for Φ, X is:

$$\mathcal{L} = \frac{1}{8\pi} \int d^4\theta \Gamma \Gamma^\dagger$$

We would like now an $N = 2$ invariant operator to provide a potential for ϕ and to set the scale. There is a very natural choice for the potential, which preserves the $N = 2$ symmetry:

$$\mathcal{L} = \frac{1}{8\pi} \int d^4\theta \Gamma^\dagger \Gamma + i \frac{\mu}{8\pi\gamma} \int d^2\theta e^{-\gamma\Gamma} + i \frac{\mu}{8\pi\gamma} \int d^2\theta^\dagger e^{-\gamma\Gamma^\dagger} \tag{3.5}$$

³ There is an analogous requirement in compactified critical string theory.

The potential is an F-term, while from the $N = 1$ point of view it is a tachyon condensate at some particular non zero momentum. This value of the momentum is special because (3.5) is actually manifestly $N = 2$ supersymmetric; in fact, it is precisely the $N = 2$ Liouville Lagrangian. However its role here is to supply an accidental global symmetry – it is *not* a remnant of $N = 2$ supergravity in superconformal gauge. The parameter γ is not renormalized in $N = 2$ Liouville (unlike in $N = 0, 1$ Liouville): $\gamma = \frac{1}{Q}$. The $N = 2$ Liouville system (3.5) together with the $2n$ X_i forms a vacuum of ($N = 1$) fermionic string theory with a global $N = 2$ symmetry. However, this is not the superstring yet – the theory still contains tachyons (3.3), does not contain space-time fermions, etc. To perform the chiral GSO projection, we define the target space SUSY operator S :

$$S(z) = e^{-\frac{1}{2}\sigma + \frac{i}{2}(H_1 + \dots + H_n + H_l + Qx)} \quad (3.6)$$

Here σ is the bosonized superconformal ghost current [28]: $\beta\gamma = \partial\sigma$; the H_i are bosonized fermions: $\psi_{2i-1}\psi_{2i} = \partial H_i$ ($i = 1, \dots, n$), and similarly $\psi_l\psi_x = \partial H_l$. The GSO projection amounts to keeping in the spectrum only operators which have a local OPE with $S(z)$. Note that apart from the last term ($e^{\frac{i}{2}Qx}$) in S , this just means projecting by $(-)^{F_R}$, where F_R is the right moving fermion number (the naive projection $\psi \rightarrow -\psi$). For $Q = 0$ (the critical case) we recover the original GSO projection. In general, the projection ties the momentum in the x direction with the fermion number. In particular, p_x takes discrete values. The operator $S(z)$ is chiral ($\bar{\partial}S = 0$); this seems strange in view of the fact that x can be non-compact (so that only operators with $p_{\text{left}} = p_{\text{right}}$ are physical). The point is that the spectrum of x momenta is determined by locality w.r.t. S ; after the projection, the radius of x is not a meaningful concept.

After choosing the “longitudinal” direction, the theory still has $SO(2n)$ rotation invariance in the X_i directions. This means that the supercharge S (3.6) is not unique. Additional charges may be obtained by acting on it with the $SO(2n)$ generators $e^{\pm iH_a \pm iH_b}$, ($a \neq b = 1, 2, \dots, n$). This gives a set of charges S_α in one of the spinor representations of $SO(2n)$. To get the other representation (for $n > 0$), one uses the fact that the operator:

$$\tilde{S}(z) = e^{-\frac{1}{2}\sigma + \frac{i}{2}(-H_1 - \dots - H_{n-1} + H_n - H_l - Qx)} \quad (3.7)$$

is local w.r.t. S , and therefore is physical. Applying $SO(2n)$ rotations to $\tilde{S}$ generates the second spinor representation. For odd n the two are isomorphic, while for even n , S_α and $S_{\tilde{\beta}}$ transform as different representations. The zero modes:

$$Q_\alpha = \oint \frac{dz}{2\pi i} S_\alpha(z); \quad \tilde{Q}_{\tilde{\beta}} = \oint \frac{dz}{2\pi i} \tilde{S}_{\tilde{\beta}}(z) \quad (3.8)$$

satisfy the super-algebra

$$\begin{aligned} \{Q_\alpha, \tilde{Q}_{\dot{\beta}}\} &= \gamma_{\alpha\dot{\beta}}^i P_i && \text{for } n \text{ even} \\ \{Q_\alpha, \tilde{Q}_{\beta}\} &= \gamma_{\alpha\beta}^i P_i && \text{for } n \text{ odd} \end{aligned} \quad (3.9)$$

The algebra (3.9) is a “space SUSY” algebra in the transverse directions: lightcone (ϕ_l, x) translations do not appear on the r.h.s., essentially because the vacuum is not translationally invariant in ϕ .

We now turn to discuss some general features of the spectrum of these theories. Although most states are paired by the SUSY algebra (3.9), apriori some tachyonic states could remain, since lack of tachyons is no longer implied by the algebra. However, for unitary matter theories (such as (3.1)), one can check directly in general that the theory is tachyon free after the GSO projection. From the point of view of section 2 this is clear: the effective density of states of the GSO projected theory (2.7) receives contributions only from states which are killed by the supercharges:

$$Q_\alpha |\text{phys}\rangle = \tilde{Q}_{\dot{\beta}} |\text{phys}\rangle = 0 \quad (3.10)$$

The rest of the states are paired level by level and cancel in the free energy (2.7). Now (3.10) implies that $|\text{phys}\rangle$ is independent of the transverse excitations (more generally of the transverse excitations and the internal degrees of freedom); the number of states that can satisfy (3.10) is therefore of the order of the number of states in the Φ, X (or Γ) system, which is a 2D string system. Hence $\lim_{s \rightarrow 0} G(s)$ is finite, and as shown in section 2, tachyons can not exist by modular invariance. Unitarity enters in two places:

1) Condition (2) of the theorem proved in section 2 breaks down in general in non unitary theories.

2) If the conditions are satisfied, the argument above proves only that the one loop partition sum is finite. There could be cancellations between NS and R tachyons, leading to a finite partition sum in the presence of tachyons. In unitary theories, Ramond tachyons can not occur (even before the chiral projection). Hence, only (NS, NS) tachyons may exist, and these are ruled out by the above argument.

Absence of tachyons can also be established directly [19], but we feel that the above argument is more intuitive. In the D dimensional superstring the explicit check of absence of tachyons is particularly simple: the only tachyon before the projection is T_k (3.3). The GSO projection (locality of T_k w.r.t. S (3.6)) implies: $k_x Q \in 2Z + 1$. The smallest value of

k_x (corresponding to the lightest state) is $|k_x| = \frac{1}{Q}$, for which $(\beta + \frac{Q}{2})^2 = \frac{Q^2}{4} + \frac{1}{Q^2} - 1 \geq 0$ for all Q . Hence the momentum in the x direction is quantized precisely such that the tachyon is always massive. In particular, the zero momentum tachyon (the original cosmological term in (3.2)) is projected out⁴.

Another interesting property of these theories is that target SUSY can be broken continuously, by switching on the modulus $\lambda \int \partial x \bar{\partial} x$. This perturbation is not projected out by GSO; it is clear from the form (3.6) of the generators that the symmetry is broken for any $\lambda \neq 0$. The situation is quite different from the ten dimensional case, where continuous breaking of SUSY can not occur [27]. The difference is due to the time dependence of the vacuum solution, but the detailed mechanism is still unclear; the gravitino is massive here even when SUSY is unbroken.

Returning to the $N = 2$ Liouville action, one can ask what is the role of the $N = 2$ cosmological term: why did we choose an $N = 2$ invariant perturbation to set the scale, out of the infinite number of GSO invariant operators, which are not $N = 2$ invariant (in general). After all, the analysis of the spectrum and symmetries are in any case performed in the region where the interaction is negligible. The answer is that it is indeed admissible to set the scale with other on shell physical vertex operators. Nevertheless, we do expect the $N = 2$ invariant perturbation (3.5) to exhibit especially nice properties. First, since (3.5) is $N = 2$ invariant, we have manifest space SUSY for all μ . For other perturbations in (3.5), SUSY would in general be broken. In addition, F-terms in (2,2) theories are known to yield moduli of the appropriate CFT (see e.g. [30]). For generic perturbations, the dynamics of the corresponding “Liouville” model (3.5) is probably much more complicated.

The $N = 2$ cosmological constant has $\beta = -\gamma = -\frac{1}{Q}$, or $E = \beta + \frac{Q}{2} = \frac{Q}{2} - \frac{1}{Q}$. This is positive only when $Q^2 > 2$. As $Q^2 \rightarrow 2$ ($d \rightarrow 5$), the lowest lying NS state goes to zero mass, and for $d > 5$, it energy becomes negative (while the mass increases). As mentioned above, the behavior of $E > 0$ and $E < 0$ states is qualitatively different. One would need a more detailed understanding of the theory to see what changes as $E \rightarrow 0$, but it is clear that the point $Q^2 = 2$ corresponds to some kind of transition in the behavior of the theory. For $Q^2 < 2$ there are other possibilities to set the scale in (3.5) with operators of positive energy, however they are all less symmetric. The special role of this new “critical dimension” is not entirely clear, and should be elucidated further.

⁴ Note also that the remaining modes of the tachyon go to infinite mass as $Q \rightarrow 0$. Therefore, they are invisible in the critical theory.

The GSO procedure may superficially resemble fine tuning the coefficients of all the tachyonic modes in the action to zero. The difference between this and fine tuning is that in the GSO procedure the states are projected out by gauge invariance (the Z_2 R symmetry). Therefore, we are assured that they will not reappear in various intermediate channels in higher loops. On the other hand, as we saw in section 2, fine tuning does not solve the problem of higher loops (topologies).

An interesting property of the non-critical superstrings we have constructed is that they have a vanishing partition sum (generically) to all orders in the genus expansion. Consider first the partition sum on the sphere: usually, the spherical partition sum is not zero in non-critical string theory [31]. The reason is the following. In critical string theory, the partition sum vanishes due to the six $c, \bar{c}$ (reparametrization ghost) zero modes on the sphere. However, in the non-critical case, the Liouville path integral contains a comparable divergence, due to zero modes of the classical Liouville solution. The zero related to the $c, \bar{c}$ zero modes is really $\frac{1}{\text{vol}SL(2,C)}$, and the infinity in the Liouville sector is proportional to the same volume (up to a finite factor), thus the two cancel. More precisely, choosing the conformal gauge does not fix the gauge completely. By fixing the $SL(2,C)$ invariance and therefore not integrating over the various zero modes, one finds a finite answer. In fermionic string theory a similar phenomenon occurs, with $SL(2,C)$ replaced by $OSp(2,1)$. Again, the infinite factors cancel, and one is left with a finite partition sum. In the limit $Q \rightarrow 0$ (when the non-critical string approaches a critical one), we still have a finite partition sum, but exactly at $Q = 0$, ϕ translation invariance appears, and we have to divide this finite answer by the volume of ϕ . The free energy per unit volume $\frac{\text{const}}{V} \rightarrow 0$ as $V \rightarrow \infty$ (ϕ is non compact). Thus, in space-time the difference between the critical and the non-critical cases is that in the latter, due to lack of ϕ translation invariance, we compute the total free energy, which is finite, while in the critical case, where translational invariance is restored, we are interested in the free energy per unit volume, which is zero.

When the fermionic string vacuum possesses a global $N = 2$ symmetry (even before the chiral projection, which is of course irrelevant for the partition sum on the sphere), there are *two additional* Liouville fermion zero modes, obtained by applying $N = 2$ transformations to the usual fermionic zero modes, which exist for $N = 1$ Liouville. The ghosts, which are still those of the $N = 1$ string do not have balancing zero modes, therefore the path integral vanishes. This means that the classical vacuum energy vanishes in (3.5). The torus, and higher genus partition sums also vanish for $n \geq 1$, by the usual contour deformation arguments [28], [32]. In that case, bosons and fermions are paired, except perhaps for

a set of measure zero of states at zero momentum $p_i = 0$ ($i = 1, \dots, 2n$). Sometimes, such arguments can be subtle due to contact terms from boundaries of moduli space, however here we expect such subtleties to be absent in general, since there are no massless excitations.

To illustrate the above abstract discussion, we finish this section with a brief analysis of the simplest theories constructed here: the two and four dimensional superstrings ($n = 0, 1$).

Example 1: Two dimensional superstring⁵.

This theory contains two superfields: the super Liouville field Φ , and a space coordinate X ; the two combine into an $N = 2$ Liouville system (3.4) with $\gamma(= \frac{1}{Q}) = \frac{1}{2}$. To calculate the torus partition sum, we have to sum (2.1) over all states satisfying:

(a) Locality w.r.t. (3.6); here, $S(z) = e^{-\frac{1}{2}\sigma + \frac{i}{2}H + iX}$.

(b) $\Delta - \bar{\Delta} \in Z$.

(c) Mutual locality.

There are four sectors in the theory. We will next go over them and solve the conditions

(a) – (c).

Consider first the (NS, NS) sector: the vertex operators have the form

$$e^{-\sigma + inH_1 + ipX} e^{-\sigma + in\bar{H}_1 + i\bar{p}\bar{X}}$$

Condition (a) leads to

$$\begin{aligned} p &= \frac{n-1}{2} + m \\ \bar{p} &= \frac{\bar{n}-1}{2} + \bar{m} \end{aligned} \tag{3.11}$$

($m, \bar{m} \in Z$). From condition (b) we have

$$\begin{aligned} n - \bar{n} &\in 2Z \\ p^2 - \bar{p}^2 &\in 2Z \end{aligned} \tag{3.12}$$

while from condition (c):

$$pp' - \bar{p}\bar{p}' \in Z \tag{3.13}$$

for all p, p' in the (NS, NS) spectrum. The solution of the constraints (3.11) – (3.13) is:

(1) $n, \bar{n} \in 2Z + 1$; $p = m$, $\bar{p} = \bar{m}$, $m - \bar{m} \in 2Z$.

⁵ Based on [33].

(2) $n, \bar{n} \in 2Z$; $p = m + \frac{1}{2}$, $\bar{p} = \bar{m} + \frac{1}{2}$, $m - \bar{m} \in 2Z$.

Summing (3.4) over these states gives (after multiplying by the oscillator contribution for X , H_I , and the ghost contribution):

$$Z_{\text{NS,NS}} = \frac{\sqrt{\tau_2}}{2} |\theta_2|^2 \quad (3.14)$$

A similar analysis applied to the other 3 sectors gives

$$Z_{\text{NS,R}} = Z_{\text{R,NS}} = -\frac{\sqrt{\tau_2}}{4} |\theta_2|^2 \quad (3.15)$$

$$Z_{\text{R,R}} = \sqrt{\tau_2} [|\theta_2|^2 + |\theta_3|^2 + |\theta_4|^2] \quad (3.16)$$

To see better where the different contributions are coming from, we should list the physical operators which survive the projection. Chirally, in the NS sector we have the tachyon operators:

$$T_k = e^{-\sigma + ikX + \beta\phi_I}; \quad k \in Z + \frac{1}{2} \quad (3.17)$$

The restriction on the spectrum of k 's is due to the requirement of locality with $S(Z)$ (3.6).

In the Ramond sector we have two sets of operators of the form:

$$V_k = e^{-\frac{1}{2}\sigma + \frac{i}{2}\epsilon H + ikX + \beta\phi_I} \quad (3.18)$$

with

$$\begin{aligned} (1) \quad & \epsilon = 1 \quad k = n = 0, 1, 2, 3, \dots; \quad \beta = n - 1 \\ (2) \quad & \epsilon = -1 \quad k = -l - \frac{1}{2} \quad l = 0, 1, 2, 3, \dots; \quad \beta = l - \frac{1}{2} \end{aligned} \quad (3.19)$$

In agreement with the partition sums (3.14) – (3.16). Due to the low space-time dimension, the superalgebra (3.9) is quite degenerate: there is only one supercharge, $Q = \oint \frac{dz}{2\pi i} S(z)$, (3.6), which satisfies $Q^2 = 0$. Hence, Q is a BRST like charge. An interesting theory is obtained by restricting ourselves to $\ker Q$: physical operators are those which satisfy:

$$Q|\text{phys}\rangle = 0 \quad (3.20)$$

This gives a topological string theory (reminiscent of the topological model of the bosonic string with $c = -2$ matter [34]). The condition (3.20) projects out the tachyon modes T_k (3.17) with $k < 0$, and the second set of Ramond states in (3.19) (those with $\epsilon = -1$, $k < 0$). Since the physical states must have the same Liouville momentum for left

and right movers (ϕ_l is non compact), all space-time fermions are projected out by (3.20). The topological two dimensional superstring contains (NS, NS) tachyons

$$T_n = e^{-\sigma + i(n + \frac{1}{2})x + (n - \frac{1}{2})\phi_l}; \quad n = 0, 1, 2, \dots \quad (3.21)$$

and (R, R) states:

$$V_n = e^{-\frac{1}{2}\sigma + \frac{1}{2}H_l + inx + (n-1)\phi_l}; \quad n = 0, 1, 2, \dots \quad (3.22)$$

(here we mean left-right symmetric combinations $\bar{V}_{\text{left}} V_{\text{right}}$). Correlation functions of the operators (3.21), (3.22) can be obtained using the methods of [35]. We leave further investigation of this topological theory to future work.

If one does not impose (3.20) on the spectrum, one can ask how does the theory change when we add the (physical) operator $\partial x \bar{\partial} x$ to the action. The dependence of the partition sum on the “radius” of x can be obtained by viewing the theory as a chiral orbifold. The idea is the following: if x has radius R ($x \simeq x + 2\pi R$), we can study the orbifold theory obtained by gauging the Z_2 symmetry:

$$\begin{aligned} x &\rightarrow x + \pi R \\ H_l &\rightarrow H_l + \pi \\ \bar{H}_l &\rightarrow \bar{H}_l \\ \sigma &\rightarrow \sigma - i\pi \\ \bar{\sigma} &\rightarrow \bar{\sigma} + 2\pi i \end{aligned} \quad (3.23)$$

This formulation of the theory as a chiral orbifold emphasizes the role of the Z_2 symmetry which is being gauged (3.23). The role of the space SUSY operator $S(z)$ is obscure from this point of view. In particular, the symmetry generator Q (3.8) is in the spectrum only for particular R ($R = 2$). This symmetric point is not singled out in (3.23).

By standard methods one obtains the partition sums:

$$Z_{NS,NS}(R) = \sqrt{\tau_2} \sum_{m, m' \in \mathbb{Z}} q^{\frac{1}{2}(\frac{2m+1}{R} + \frac{m'R}{2})^2} \bar{q}^{\frac{1}{2}(\frac{2m+1}{R} - \frac{m'R}{2})^2} \quad (3.24)$$

$$Z_{R,R}(R) = \sqrt{\tau_2} \sum_{m, m' \in \mathbb{Z}} q^{\frac{1}{2}(\frac{m}{R} + \frac{m'R}{2})^2} \bar{q}^{\frac{1}{2}(\frac{m}{R} - \frac{m'R}{2})^2} \quad (3.25)$$

$$Z_{NS,R}(R) + Z_{R,NS}(R) = -\sqrt{\tau_2} \sum_{m, m' \in \mathbb{Z}} q^{\frac{1}{2}(\frac{2m}{R} + \frac{R}{2}(m' + \frac{1}{2}))^2} \bar{q}^{\frac{1}{2}(\frac{2m}{R} - \frac{R}{2}(m' + \frac{1}{2}))^2} \quad (3.26)$$

For $R = 2$ we have the results (3.14), (3.15), (3.16). For generic R , the total partition sum can be written as a linear combination of $c = 1$ partition sums: $Z_{\text{total}}(R) = 2Z_{c=1}(R) - Z_{c=1}(\frac{4}{R})$. The integral $\Omega = \int_{\mathcal{F}} \frac{d^2\tau}{\tau^2} Z_{\text{total}}$, which as we have learnt in section two gives the number of degrees of freedom, can be performed using (2.11) (or the results of [36]); one finds that $\Omega \propto R$. In particular, at the “supersymmetric”, or topological, point $R = 2$, the cancellation of bosons and fermions is not complete. This is due to the low space-time dimension – the SUSY algebra $Q^2 = 0$ does not imply pairing of bosons and fermions. As explained above, we need at least two non compact transverse directions for that. Therefore, we will next consider the case of two transverse dimensions.

Example 2: Four dimensional superstring.

The “matter” consists of two superfields X_1, X_2 (in addition to Φ, X), $Q = \sqrt{3}$, and $S = e^{-\frac{1}{2}\sigma + \frac{1}{2}(H_1 + H_2 + \sqrt{3}x)}$. One can repeat the discussion of Example 1 here; we will leave the details to the reader, and only give the final result for the one loop partition sums⁶:

$$\begin{aligned}
 \sqrt{\tau_2} |\eta(\tau)|^6 Z_{NS, NS} = & |\theta_3 + \theta_4|^2 \sum_{\epsilon} \Theta \left[\begin{smallmatrix} \frac{\epsilon+3}{6} \\ 0 \end{smallmatrix} \right] (12\tau) \bar{\Theta} \left[\begin{smallmatrix} \frac{\epsilon+3}{6} \\ 0 \end{smallmatrix} \right] (12\bar{\tau}) + \\
 & |\theta_3 - \theta_4|^2 \sum_{\epsilon} \Theta \left[\begin{smallmatrix} \frac{\epsilon}{6} \\ 0 \end{smallmatrix} \right] (12\tau) \bar{\Theta} \left[\begin{smallmatrix} \frac{\epsilon}{6} \\ 0 \end{smallmatrix} \right] (12\bar{\tau}) + \\
 & (\theta_3 + \theta_4)(\bar{\theta}_3 - \bar{\theta}_4) \sum_{\epsilon} \Theta \left[\begin{smallmatrix} \frac{\epsilon+3}{6} \\ 0 \end{smallmatrix} \right] (12\tau) \bar{\Theta} \left[\begin{smallmatrix} \frac{\epsilon}{6} \\ 0 \end{smallmatrix} \right] (12\bar{\tau}) + \\
 & (\theta_3 - \theta_4)(\bar{\theta}_3 + \bar{\theta}_4) \sum_{\epsilon} \Theta \left[\begin{smallmatrix} \frac{\epsilon}{6} \\ 0 \end{smallmatrix} \right] (12\tau) \bar{\Theta} \left[\begin{smallmatrix} \frac{\epsilon+3}{6} \\ 0 \end{smallmatrix} \right] (12\bar{\tau})
 \end{aligned} \tag{3.27}$$

$$\begin{aligned}
 \sqrt{\tau_2} |\eta(\tau)|^6 Z_{R, R} = & |\theta_2|^2 \sum_{\epsilon} \Theta \left[\begin{smallmatrix} \frac{\epsilon-\frac{1}{2}}{6} \\ 0 \end{smallmatrix} \right] (12\tau) \bar{\Theta} \left[\begin{smallmatrix} \frac{\epsilon-\frac{1}{2}}{6} \\ 0 \end{smallmatrix} \right] (12\bar{\tau}) + \\
 & |\theta_2|^2 \sum_{\epsilon} \Theta \left[\begin{smallmatrix} \frac{\epsilon+3-\frac{1}{2}}{6} \\ 0 \end{smallmatrix} \right] (12\tau) \bar{\Theta} \left[\begin{smallmatrix} \frac{\epsilon+3-\frac{1}{2}}{6} \\ 0 \end{smallmatrix} \right] (12\bar{\tau}) + \\
 & |\theta_2|^2 \sum_{\epsilon} \Theta \left[\begin{smallmatrix} \frac{\epsilon-\frac{1}{2}}{6} \\ 0 \end{smallmatrix} \right] (12\tau) \bar{\Theta} \left[\begin{smallmatrix} \frac{\epsilon+3-\frac{1}{2}}{6} \\ 0 \end{smallmatrix} \right] (12\bar{\tau}) + \\
 & |\theta_2|^2 \sum_{\epsilon} \Theta \left[\begin{smallmatrix} \frac{\epsilon+3-\frac{1}{2}}{6} \\ 0 \end{smallmatrix} \right] (12\tau) \bar{\Theta} \left[\begin{smallmatrix} \frac{\epsilon-\frac{1}{2}}{6} \\ 0 \end{smallmatrix} \right] (12\bar{\tau}) +
 \end{aligned} \tag{3.28}$$

⁶ All sums over ϵ run over $\epsilon = 0, 2, 4$.

$$\begin{aligned}
-\sqrt{\tau_2}|\eta(\tau)|^6 Z_{NS,R} = & (\theta_3 + \theta_4)\bar{\theta}_2 \sum_{\epsilon} \Theta \left[\frac{\frac{\epsilon+3}{6}}{0} \right] (12\tau) \bar{\Theta} \left[\frac{\frac{\epsilon+\frac{3}{2}}{6}}{0} \right] (12\bar{\tau}) + \\
& (\theta_3 - \theta_4)\bar{\theta}_2 \sum_{\epsilon} \Theta \left[\frac{\frac{\epsilon}{6}}{0} \right] (12\tau) \bar{\Theta} \left[\frac{\frac{\epsilon-\frac{3}{2}}{6}}{0} \right] (12\bar{\tau}) + \\
& (\theta_3 + \theta_4)\bar{\theta}_2 \sum_{\epsilon} \Theta \left[\frac{\frac{\epsilon+3}{6}}{0} \right] (12\tau) \bar{\Theta} \left[\frac{\frac{\epsilon-\frac{3}{2}}{6}}{0} \right] (12\bar{\tau}) + \\
& (\theta_3 - \theta_4)\bar{\theta}_2 \sum_{\epsilon} \Theta \left[\frac{\frac{\epsilon}{6}}{0} \right] (12\tau) \bar{\Theta} \left[\frac{\frac{\epsilon+\frac{3}{2}}{6}}{0} \right] (12\bar{\tau}) +
\end{aligned} \tag{3.29}$$

Now, by general arguments above, we expect $Z_{\text{total}} = Z_{NS,NS} + Z_{R,R} + Z_{NS,R} + Z_{R,NS}$ to vanish! Quite miraculously from the point of view of the explicit formulae (3.27) – (3.29), one can write Z_{total} in a much simpler way as:

$$\sqrt{\tau_2}|\eta(\tau)|^6 Z_{\text{total}} = \sum_{\epsilon} |F_{\epsilon}(\tau)|^2 \tag{3.30}$$

where

$$\begin{aligned}
F_{\epsilon}(\tau) = & (\theta_3 + \theta_4)(\tau) \Theta \left[\frac{\frac{\epsilon+3}{6}}{0} \right] (12\tau) + (\theta_3 - \theta_4)(\tau) \Theta \left[\frac{\frac{\epsilon}{6}}{0} \right] (12\tau) \\
& - \theta_2(\tau) \left\{ \Theta \left[\frac{\frac{\epsilon+\frac{3}{2}}{6}}{0} \right] (12\tau) + \Theta \left[\frac{\frac{\epsilon-\frac{3}{2}}{6}}{0} \right] (12\tau) \right\}
\end{aligned} \tag{3.31}$$

The most natural scenario would be to have $F_{\epsilon}(\tau) = 0$ for all $\epsilon (= 0, 2, 4)$. This seems indeed to be the case. We have checked the vanishing of F_{ϵ} (3.31) for low orders in q , and believe that the result is true in general, although we haven't proved it analytically. If this is the case, we have, as expected, $Z_{\text{total}} = 0$.

4. CLASSICAL DYNAMICS IN 2D STRING THEORY.

In the previous sections we have seen that stable (tachyon free) string vacua are in general those with $c_{\text{string}} = 1$. Bosonic strings with $c_{\text{string}} = 1$ (two dimensional strings), are known from the matrix models [5], [6], [37], [38] to be solvable. In this section we will review the understanding of their properties in the continuum path integral formalism, in the hope that some of them may carry over and help to understand more complicated theories with $c_{\text{string}} = 1$, like those of section 3.

In the introduction, we have seen that the set of “resonant” amplitudes defines for all d a “bulk” S – matrix, which is given by Shapiro – Virasoro type integral representations.

To illustrate this, we discuss here tachyon dynamics in $D = d+1$ dimensional “non-critical” string theory. Of course, for generic D , there is no reason to concentrate on the tachyon, both because it is merely the lowest lying state of the string spectrum, and because it is tachyonic, thus absent in more physical theories (see section 3). Our justification is that for $D = 2$ it is the only field theoretic degree of freedom, and is massless.

The on shell vertex operator for the tachyon is:

$$T_k = \exp(ik \cdot X + \beta(k)\phi) \quad (4.1)$$

where k, X are d - vectors, and by (1.5):

$$\frac{1}{2}k^2 - \frac{1}{2}\beta(\beta + Q) = 1 \quad (4.2)$$

which implies as before (1.8): $m_{\text{tachyon}}^2 = \frac{2-D}{12}$. The bulk amplitudes described in the introduction take here the form:

$$A_{s=0}(k_1, \dots, k_N) = \langle T_{k_1} \dots T_{k_N} \rangle / \log \mu \quad (4.3)$$

where $\sum_i k_i = 0$, and by (1.9) $\sum_i \beta_i = -Q$. As mentioned above, such amplitudes have an integral representation for all D :

$$A_{s=0}(k_1, \dots, k_N) = \prod_{i=4}^N \int d^2 z_i |z_i|^{2(k_1 \cdot k_i - \beta_1 \beta_i)} |1 - z_i|^{2(k_3 \cdot k_i - \beta_3 \beta_i)} \prod_{4=i < j}^N |z_i - z_j|^{2(k_i \cdot k_j - \beta_i \beta_j)} \quad (4.4)$$

For $D = 26$ this is the familiar Shapiro - Virasoro representation [1], however it retains all the nice properties for all D . The singularities, which are the most important features of (4.4), are poles in various channels, corresponding to on shell intermediate particles. To examine these poles, one focuses on contributions to (4.4) of regions in moduli space where vertices approach each other. E.g. from the region $z_4, z_5, \dots, z_{n+2} \rightarrow 0$, we have an infinite series of poles in (E, p) , $E = \frac{Q}{2} + \sum_i \beta_i$, $p = \sum_i k_i$ ($i = 1, 4, 5, 6, \dots, n+2$), with residues related to lower correlation functions with insertions of various intermediate states. For example, near the first (tachyon) pole we have:

$$A_{s=0}(k_1, \dots, k_N) \simeq \frac{\langle T_{k_1} T_{k_4} \dots T_{k_{n+2}} T_{\tilde{k}} \rangle \langle T_{\Sigma_i k_i} T_{k_2} T_{k_3} T_{k_{n+3}} \dots T_{k_N} \rangle}{(\frac{Q}{2} + \sum_i \beta_i)^2 - (\sum_i k_i)^2 - \frac{2-D}{12}} \quad (4.5)$$

where $\vec{k} = -\sum_i k_i$, and similarly for the higher poles. Thus, for generic correlators (4.4), the pole structure is very complicated. For $N = 4$, a closed expression for (4.4) in terms of Γ functions is known [1]:

$$A_{s=0}(k_1, \dots, k_4) = \pi \prod_{i=2}^4 \frac{\Gamma(k_1 \cdot k_i - \beta_1 \beta_i + 1)}{\Gamma(\beta_1 \beta_i - k_1 \cdot k_i)} \quad (4.6)$$

It is easy to check that (4.6) satisfies factorization (4.5). For $N \geq 5$ no such closed forms are known.

The situation is markedly different in two dimensions. There, the singularity structure of (4.4) is much simpler, and the integral representation can be evaluated for all N . The condition (4.2) implies that we have left and right moving massless “tachyons” in space-time:

$$T_k^{(\pm)} = \exp(ikX + \beta^{(\pm)}\phi); \quad \beta^{(\pm)} = -\sqrt{2} \pm k \quad (4.7)$$

The correlation functions have then the following form [39]:

$$\begin{aligned} \langle T_{k_1}^{(+)} \dots T_{k_{N-1}}^{(+)} T_{k_N}^{(-)} \rangle &= (N-3)! \prod_{i=1}^N (-\pi) \frac{\Gamma(\frac{1}{2}\beta_i^2 - \frac{1}{2}k_i^2)}{\Gamma(1 - \frac{1}{2}\beta_i^2 + \frac{1}{2}k_i^2)} \\ \langle T_{k_1}^{(+)} \dots T_{k_n}^{(+)} T_{p_1}^{(-)} \dots T_{p_m}^{(-)} \rangle &= 0; \quad n, m \geq 2 \end{aligned} \quad (4.8)$$

The Γ factor for $i = N$ in (4.8) is infinite. This infinity should be interpreted as the volume of the Liouville mode $\log \mu$ (see section 1). The vanishing of the second set of correlators in (4.8) can be understood as a lack of such a volume factor. Taking this into account, the first form properly understood describes all bulk N point functions [39], [35].

The results (4.8) are at first sight very puzzling. Both the vanishing of amplitudes with $n, m \geq 2$, and the simple form of the amplitudes with $m = 1$ (or $n = 1$) are not at all obvious from (4.4); apriori one would expect *much more* poles, as in (4.5). We will not derive (4.8) here (the reader may find a more complete discussion in [39], [35]), but we will explain the origin of its simple form. The phenomenon behind this simplicity is partial decoupling of a certain infinite set of discrete states. To understand why discrete states are important, we have to recall the spectrum of the two dimensional string [40], [22], [41]. We have already encountered the tachyon field (4.7). For $D > 2$ there is in addition an infinite tower of massive transverse oscillator states. For $D = 2$ there are no transverse directions, hence most of the massive degrees of freedom are absent. There are still oscillator states at discrete values of the momentum, which are related to oscillator

primary states of $c = 1$ CFT [42] (in space-time these are longitudinal oscillations of the string). The general form of these states is:

$$V_{r_1, r_2}^{(\pm)} = [\partial^{r_1 r_2} X + \dots] \exp \left(i \frac{r_1 - r_2}{\sqrt{2}} X + \beta_{r_1, r_2}^{(\pm)} \phi \right) \quad (4.9)$$

where $r_1, r_2 \in \mathbb{Z}_+$, and the $\dots$ stands for a certain polynomial in $\partial X, \partial^2 X, \dots$ (with scaling dimension $r_1 r_2$). The Liouville dressing is $\beta_{r_1, r_2}^{(\pm)} = -\sqrt{2} \pm \frac{r_1 + r_2}{\sqrt{2}}$; note that unlike (4.7), here $V^{(+)}$ ($V^{(-)}$) correspond to positive (negative) energy states.

The states (4.9) together with the tachyon comprise the full spectrum of 2d string theory (at ghost number zero). Thus, all N point functions (4.4) must factorize on these states. To demonstrate the role of the discrete states in (4.4), we should analyze the factorization in various channels. Consider, for example, the simplest degeneration channel, when two tachyons collide. There are two possibilities (up to $X \rightarrow -X$): 1) $T_{k_1}^{(+)}(z)T_{k_2}^{(-)}(0)$; 2) $T_{k_1}^{(+)}(z)T_{k_2}^{(+)}(0)$, $z \rightarrow 0$.

In the first case, using the free OPE, which is clearly valid in bulk correlators, although not in general [8], [20], we find an infinite set of poles at $k_1 k_2 + (k_1 - \sqrt{2})(k_2 + \sqrt{2}) = -n$, $n = 1, 2, 3, \dots$. The first pole ($n = 1$) corresponds to an intermediate tachyon, while the rest must have vanishing residues. The reason is that the residue involves correlators of oscillator states (generalizing (4.5)), but the momentum $k_1 + k_2$ of the intermediate state is *not* in the discrete list (4.9) (at least for generic k_1, k_2), therefore the corresponding state must be BRST trivial (null), and decouple.

In the second case, the poles occur at $k_1 k_2 - (k_1 - \sqrt{2})(k_2 - \sqrt{2}) = -n$ or: $\sqrt{2}(k_1 + k_2) = 2 - n$. The intermediate momentum is automatically in the discrete list (4.9); furthermore $\beta = \beta_1 + \beta_2 = -\sqrt{2} - \frac{n}{\sqrt{2}}$ is such that the energy of the intermediate state is negative – it is a $V^{(-)}$. A similar situation occurs for poles in more complicated channels.

Therefore, the pole structure of (4.4) depends crucially on the physics of the states $V_{r_1, r_2}^{(-)}$. All but a finite number of the poles in the tachyon S -matrix correspond to these states. One can show [35] that they decouple from amplitudes of tachyons,

$$\langle V_{r_1, r_2}^{(-)} T_{k_1} \dots T_{k_N} \rangle = 0 \quad (4.10)$$

as long as $k_1, \dots, k_N \notin \frac{1}{\sqrt{2}}\mathbb{Z}$. Therefore, most of the poles in (4.4) have vanishing residues. Using (4.10) and factorization it is straightforward to show that most of the tachyon amplitudes vanish, and the ones with signatures $(N - 1, 1)$ have the simple structure (4.8) (see [35]).

The poles in (4.8) appear as “external leg factors” [22], however they also have a standard interpretation as poles in the (i, N) channels (i.e. when T_{k_i} and T_{k_N} approach each other) – the only channels where the poles have finite residues. Since k_N is fixed by kinematics ($k_N = -\frac{N-2}{\sqrt{2}}$), they appear to be “leg poles”. One could ask at this point why these poles have a finite residue; after all, the intermediate states are again $V^{(-)}$ in this case. The point is that the decoupling (4.10) of the discrete states with negative energies is *only partial*. If there are enough $V^{(+)}$ ’s, the amplitude does not vanish. The residues of the poles in the (i, N) channel involve the three point function $\langle V_{r_1, r_2}^{(-)} T_{k_1} T_{k_2} \rangle$. By the conservation laws, k_1, k_2 are in this case in the discrete list (4.9) and their energies are positive; this correlation function does not vanish. The general situation is the following: the most general correlator has the form:

$$\langle V_{r_1, r_2}^{(+)} \dots V_{r_{2n-1}, r_{2n}}^{(+)} V_{s_1, s_2}^{(-)} \dots V_{s_{2l-1}, s_{2l}}^{(-)} T_{k_1} \dots T_{k_N} \rangle \propto (\log \mu)^{n-l} \quad (4.11)$$

where $k_1, \dots, k_N \notin \frac{1}{\sqrt{2}}\mathbb{Z}$. Eq. (4.11) is very useful to determine the relative size of various amplitudes. If the power of $\log \mu$ ($n-l$) is one, we have a finite bulk amplitude. This is the case for (4.3) with signature $(N-1, 1)$ and generic $k_1, \dots, k_{N-1}$; k_N is fixed by kinematics to be $V_{0, N-2}^{(+)}$. If the power $n-l \leq 0$, the bulk amplitude (coefficient of $\log \mu$) vanishes. This is the case for tachyon amplitudes with signature (n, m) , $n, m \geq 2$, and generic momenta (here the conservation laws do not constrain individual momenta), and for correlation functions like (4.10).

One natural question that arises at this point is whether we can reconstruct the general (non bulk) amplitudes in the theory from the knowledge of the bulk ones. In general, this issue is not understood (although it is probably possible to do this). However, in 2d (closed) string theory it appears that one can make an educated guess for the answer, based on a physical assumption. Since the results obtained by this method are in agreement with those obtained from matrix models, the physical assumption is probably justified, however it has not been derived from first principles (in the continuum formalism). The idea is the following: we saw before how the structure of 2d string theory leads to the bulk correlators (4.8). It is very natural to define “renormalized operators” ($\Delta(x) \equiv \frac{\Gamma(x)}{\Gamma(1-x)}$):

$$\tilde{T}_k = \frac{T_k}{(-\pi)\Delta(\frac{1}{2}\beta^2 - \frac{1}{2}k^2)} \quad (4.12)$$

These have very simple correlation functions:

$$\langle \tilde{T}_{k_1} \dots \tilde{T}_{k_N} \rangle = (N-3)! \quad (4.13)$$

What happens for non zero s (1.9)? Integer $s > 0$ is easy to treat since if $\langle \tilde{T}_{k_1} \dots T_{k_N} \rangle = \tilde{A}_{k_1 \dots k_N} \mu^s$, then differentiating s times w.r.t. μ we find

$$\tilde{A}(k_1, \dots, k_N) = \frac{\langle \tilde{T}_{k_1} \dots \tilde{T}_{k_N} (\tilde{T}_{k=0})^s \rangle}{s!}$$

The numerator is a bulk amplitude, so we can use (4.13) and find: $A(k_1, \dots, k_N) = \frac{(s+N-3)!}{s!} \mu^s$, or:

$$A_s(k_1, \dots, k_N) = \left(\frac{\partial}{\partial \mu} \right)^{N-3} \mu^{s+N-3} \quad (4.14)$$

We can further redefine the operators (and rescale the path integral) such that all μ dependence on the r.h.s. of (4.14) disappears; alternatively put $\mu = 1$. Then we find ourselves in a situation where all correlators at integer s are polynomials in $k_1, \dots, k_N$ (using (1.9)). The question is how we can continue to non integer s . The first guess would be to declare (4.14) valid for any s – it certainly makes sense for all amplitudes. While this is true in a certain region in momentum space, in general the answer is more subtle.

There are several clues pointing in the direction we should choose. The first clue is the fact that the effect of the massive modes of the string is summarized, at least for bulk amplitudes, in the external leg factors (4.8). The renormalized field $\tilde{T}_k$ (4.12) is completely insensitive to the massive modes. This suggests that the S – matrix of $\tilde{T}$ is described by a local two dimensional field theory. One might have expected to see tachyon poles in amplitudes, however as we saw, they are not there (4.8). The reason is lack of conservation of Liouville momentum in this field theory (1.9). From the point of view of Liouville theory, we have the OPE:

$$\exp(\beta_1 \phi) \exp(\beta_2 \phi) = \int d\beta \exp(\beta \phi) f(\beta, \beta_1, \beta_2) \quad (4.15)$$

where f is an OPE coefficient. The contour of integration for β should [8] run over macroscopic states $\beta = -\frac{Q}{2} + ip$ ($p \in R$). Using the OPE (4.15) in generic tachyon correlation functions we find that the contribution of regions of moduli space where vertices approach each other is:

$$\begin{aligned} \int d^2 z \langle T_{k_1}(z) T_{k_2}(0) \dots \rangle &\simeq \int_{|z| < \epsilon} d^2 z \int_{-\infty}^{\infty} dp (z \bar{z})^{\frac{1}{2} p^2 + \frac{1}{2} (k_1 + k_2)^2 - 1} f(p, \beta_1, \beta_2) \langle T_{k_1 + k_2} \dots \rangle \\ &\simeq \int_{-\infty}^{\infty} dp \frac{f(p, \beta_1, \beta_2)}{p^2 + (k_1 + k_2)^2} \langle T_{k_1 + k_2} \dots \rangle \end{aligned} \quad (4.16)$$

where we interchanged the order of the z and p integrations. For fixed p , (4.16) has the familiar form from critical string theory; we find a pole corresponding to the intermediate state $T_{k_1+k_2}$. The fact that Liouville momentum is not conserved and we have to sum over all p 's may turn this pole into a cut: (4.16) depends on $|k_1 + k_2|$. Thus we expect cuts whenever some of the momenta $\{k_i\}$ in (4.3) satisfy $\sum_i k_i \equiv p \rightarrow 0$. The fact that the bulk $\tilde{T}$ correlators are polynomial in k_i (4.14) is another indication that $\tilde{T}$ is described by a local 2d field theory. However, due to the expected appearance of cuts in amplitudes, we have to be careful in continuing (4.14) from its region of validity. The remaining issue is to understand the expected analytic structure of the correlation functions.

First, if the picture we have been developing is correct, we expect the three point functions $\langle T_{k_1} T_{k_2} T_{k_3} \rangle$ to be given exactly by (4.14):

$$\langle \tilde{T}_{k_1} \tilde{T}_{k_2} \tilde{T}_{k_3} \rangle = 1 \quad (4.17)$$

The reason is that non analytic effects such as (4.16) can only occur for $N \geq 4$ point functions – three point functions are insensitive to the non-conservation of energy (4.15). This immediately allows us to obtain the propagator of the tachyon. By putting $k_2 = 0$ in (4.17) and integrating we obtain the two point function⁷:

$$\langle \tilde{T}_k \tilde{T}_{-k} \rangle = \frac{1}{\sqrt{2}|k|} \quad (4.18)$$

Hence the propagator (in a convenient normalization) should be: $\frac{|k|}{\sqrt{2}}$. An important consistency check on this is the comparison of an amplitude with an insertion of a puncture $P = T_{k=0}$ to the amplitude without it. By KPZ scaling (1.9) we have:

$$\langle P T_{k_1} \dots T_{k_N} \rangle = \left[\frac{1}{\sqrt{2}} \sum_{i=1}^N |k_i| - (N-2) \right] \langle T_{k_1} \dots T_{k_N} \rangle \quad (4.19)$$

Thinking of (4.19) as a relation between tree amplitudes in the purported space time field theory reveals its essential features: we can insert the puncture $T_{k=0}$ into the tree amplitude $\langle T_{k_1} \dots T_{k_N} \rangle$ either by attaching it to one of the N external legs, thus adding

⁷ In critical string theory the two point function vanishes, however, as explained above for the partition sum, this is due to a division by the (infinite) volume of ϕ . Here, due to lack of translational invariance, the two point function is finite, and is related to the propagator.

an internal propagator of momentum $k_i + 0 = k_i$ or inside the diagram. The first term (the sum) on the r.h.s. of (4.19) corresponds to the first possibility; we can read off the propagator $\frac{|k|}{\sqrt{2}}$, which agrees with that obtained from (4.18). The second term corresponds to the second possibility. Now, after understanding the propagator, the remaining problem is specifying the vertices in the space-time field theory. The three point vertex is 1 (4.17). The higher vertices can be calculated using two basic properties:

- a) The form (4.14) is exact for $k_1, \dots, k_{N-1} > 0$, $k_N < 0$, which is the region in which all the energies involved are positive (4.3); from the world sheet point of view this is natural since positive energy perturbations correspond to small deformations of the surface [8], and also because the integrals (4.4) converge there (after a certain analytic continuation in the central charge [35]).
- b) All higher irreducible vertices are *analytic* in $\{k_i\}$. To understand this, consider for example the four point function:

$$\tilde{A}(k_1, \dots, k_4) = \int d^2 z \langle \tilde{T}_{k_1}(0) \tilde{T}_{k_2}(\infty) \tilde{T}_{k_3}(1) \tilde{T}_{k_4}(z) \rangle \quad (4.20)$$

We can separate the z integral in (4.20) into two pieces. One is a sum of three contributions from the regions $z \rightarrow 0, 1, \infty$. By (4.16) we expect to get from those the tachyon propagator⁸ $\frac{1}{\sqrt{2}}|k_1 + k_i|$. The rest of the z integral is the contribution of the bulk of moduli space (where z is not close to $0, 1, \infty$); it gives a new irreducible four particle interaction (which we will denote by $A_{1PI}^{(4)}$) for the tachyons. This contribution is analytic, since only massless intermediate states cause cuts in the amplitudes (4.16), and we have subtracted their contribution. Using this decomposition, we expect the following analytic structure for $\tilde{A}(k_1, \dots, k_4)$:

$$\tilde{A}(k_1, \dots, k_4) = \left(\frac{|k_1 + k_2|}{\sqrt{2}} + \frac{|k_1 + k_3|}{\sqrt{2}} + \frac{|k_1 + k_4|}{\sqrt{2}} \right) + A_{1PI}^{(4)} \quad (4.21)$$

where $A_{1PI}^{(4)}$ is analytic in k_i . Next we use the fact that in the region $k_1, k_2, k_3 > 0$, $k_4 < 0$, we actually know $\tilde{A}$ (4.14). Comparing to (4.21) we find

$$A_{1PI}^{(4)} = -1 \quad (4.22)$$

Eq. (4.22) holds now everywhere in $\{k_i\}$ by analytic continuation.

⁸ The contribution of the massive states in the OPE (4.16) is presumably related to the factors $\Delta(\frac{1}{2}\beta_i^2 - \frac{1}{2}k_i^2)$ in (4.12).

It is now clear how to proceed in the case of N point functions. We assume that we know already $A_{1PI}^{(4)}, \dots, A_{1PI}^{(N-1)}$. Then we write all possible tree graphs with N external legs, propagator $\frac{|k|}{\sqrt{2}}$ and vertices $A_{1PI}^{(n)}$ ($n \leq N-1$) and add an unknown new irreducible vertex $A_{1PI}^{(N)}(k_1, \dots, k_N)$. $A_{1PI}^{(N)}$ is again analytic in $\{k_i\}$ and we can fix it by comparing the sum of exchange amplitudes (reducible graphs) and $A_{1PI}^{(N)}$ to the full answer (4.14) in the appropriate kinematic region. This fixes $A_{1PI}^{(N)}$ in the above kinematic region. Now we use analyticity of $A_{1PI}^{(N)}$ to fix it everywhere. The outcome of this process is the determination of the amplitudes in all kinematic regions given their values in one kinematic region.

In practice, it is more convenient to obtain $A_{1PI}^{(N)}$ by Legendre transforming. This gives a highly non trivial set of irreducible vertices; the general formulae are quite involved [35]. As an example, one finds:

$$A_{1PI}^{(N)}(k_1, k_2, k_3, k_4 = 0, \dots, k_N = 0) = (\partial_\mu)^{N-3} \left\{ \frac{1}{\mu} \prod_{i=1}^3 \frac{1}{\cosh(\frac{k_i}{\sqrt{2}} \log \mu)} \right\} \Big|_{\mu=1} \quad (4.23)$$

$$A_{1PI}^{(N)}(k_1, k_2, k_3, k_4, 0, \dots, 0) = (\partial_\mu)^{N-4} \mu^{-2} \left(\prod_{i=1}^4 \frac{1}{\cosh(\frac{k_i}{\sqrt{2}} \log \mu)} \right) \left(-1 - \mu \partial_\mu \log \prod_{1 \leq i < j \leq 3} \cosh(\frac{k_i + k_j}{\sqrt{2}} \log \mu) \right) \Big|_{\mu=1} \quad (4.24)$$

As expected, the irreducible vertices are analytic in $\{k_i\}$. This is true in general. Note that the preceeding discussion is very reminiscent of the decomposition of moduli space which appears in general in closed string field theory [43]. It would be interesting to make the relation more precise. The tachyon field theory constructed above is also closely related to the one which arises in matrix models [44], [45], [46]).

This concludes our presentation of tree level scattering in 2d closed string theory. While we now know the structure of the S - matrix in great detail, our understanding of the origin of the factorization of poles (4.8) and the emergence of the local tachyon field theory is still unsatisfactory. For example, does the spectrum (massless tachyon + discrete states) automatically lead to the structure we have described? In particular, is the decoupling of the discrete states with $E < 0$ generic?

To gain more information about the possible structures in 2d string theory, we have studied other string models with similar general characteristics. The first is the 2d fermionic string. For $N = 1$ supergravity [35], the space-time dynamics includes two massless scalars and a set of discrete states. The classical S - matrix is extremely similar

to the bosonic case: the discrete states with $E \leq 0$ again (partially) decouple, and the bulk S - matrix again has only factorized poles, as in (4.8).

The 2d $N = 2$ (critical) string was studied in [47], and its tree level S - matrix is even simpler (all $N \geq 4$ point functions vanish). This is due to the fact that the theory is really four dimensional. Open 2d strings exhibit quite different behavior from the closed case [48]. The spectrum contains again the massless field (4.7) and discrete states (4.9), however the S - matrix of the open string tachyon $T_k^{(B)}$ on the disk has the form ($m_i = \frac{1}{2}\beta_i^2 - \frac{1}{2}k_i^2$):

$$A_{\text{open}}^{(n,m)} = \left[\prod_{i=1}^{n+m} \frac{1}{\Gamma(1-m_i)} \right] H_n(k_1, \dots, k_n) H_m(p_1, \dots, p_m) \quad (4.25)$$

where

$$H_n(k_1, k_2, \dots, k_n) = \sum_{\sigma} \prod_{i=1}^{n-1} \frac{1}{\sin \pi \sum_{j=1}^i m_{\sigma(j)}} \quad (4.26)$$

and the sum over σ in (4.26) runs over permutations of $1, 2, \dots, n$. The form (4.25), (4.26) exhibits in general a rich pole structure which is in essence due to the fact that the states $V_{r_1, r_2}^{(-)}$ do not decouple here. Therefore, the decoupling of the massive states from the tachyon dynamics also doesn't occur here. Remarkably, the complicated S - matrix (4.25), (4.26) again follows from a very simple space-time field theory. Indeed, one can check [48] that the generating functional:

$$\mathcal{W}(T^{(\pm)}) = \int D\psi D\psi^* D\bar{\psi} D\bar{\psi}^* \exp \left[- \int dX d\phi e^{\frac{Q\phi}{2}} \mathcal{L}(\psi, T) \right] \quad (4.27)$$

where the Lagrangian $\mathcal{L}$ is given by:

$$\mathcal{L}(\psi, T) = \psi^* \sin(\pi\alpha_- \partial^+) \psi + \bar{\psi}^* \sin(\pi\alpha_+ \partial^-) \bar{\psi} + T^{(+)} \bar{\psi}^* \bar{\psi} + T^{(-)} \psi^* \psi, \quad (4.28)$$

and $\partial^{\pm} = \partial_{\phi} \pm \partial_X$. satisfies:

$$\frac{\delta^{n+m} \mathcal{W}}{\delta T_{k_1}^{(+)} \dots \delta T_{k_n}^{(+)} \delta T_{p_1}^{(-)} \dots \delta T_{p_m}^{(-)}} = H_n(k_1, \dots, k_n) H_m(p_1, \dots, p_m) \quad (4.29)$$

The reader should consult [48] for details and a more precise formulation. Hence, (4.27) generates the scattering amplitudes (4.25), (4.26). Furthermore, the propagator in (4.28) is essentially a lattice propagator, and it is natural to define the theory on a space-time lattice in $\frac{1}{2}(X \pm i\phi)$. Thus, the structure arising in the open sector is both richer (4.25), (4.26) and simpler (4.27), (4.28) than the one observed in the closed sector of the theory.

5. COMMENTS.

Despite the recent progress, there are still many unresolved fundamental questions both in 2d string theory, and in its relation to higher dimensional models. We have identified one feature which 2d string theory shares with higher dimensional stable theories – the fact that $c_{\text{string}} = 1$ (the small effective number of states). It is not clear what other features of 2d string theory have higher dimensional analogs. The non critical superstring models of section 3 may be useful to study this issue.

Even within the framework of 2d string theory, the understanding of the free fermion structure of [5], [6] is very incomplete. This seems to boil down to a better understanding of the dynamics of $V_{r,s}^{(\pm)}$ (4.9). Understanding the dynamics of the discrete states is crucial also to study gravitational physics in 2d string theory. The black hole of [25] is described [49] by turning on $V_{1,1}^{(-)}$; other $V_{r,s}^{(-)}$ seem also to play a role in gravitational back reaction.

Two dimensional closed string theory seems to be described by two consistent (but different) S – matrices. The S – matrix for T_k (4.1), given by (4.4) is sensitive to the discrete states (4.9), and contains the information on gravitational physics. Its pole structure and the effective action which describes it may be used to study the space-time 2d gravity. On the other hand, $\tilde{T}_k$ (4.12) are described by a second S – matrix, which originates from a two dimensional field theory for the tachyon field. In particular, it doesn't contain space-time gravity. The appearance of the local tachyon field theory is quite surprising from the point of view of the continuum formulation (as is the case with many other features of these models, it arises more naturally in the matrix model [44], [45], [46]); it would be interesting to understand this dual structure better.

A very important question concerns the form of the exact classical equations of motion in 2d string theory. The correlation functions we have found suggest a very interesting structure. In the sigma model approach, one can investigate the moduli space of classical solutions by requiring that adding perturbations to the world action: $\mathcal{S} \rightarrow \mathcal{S} + \lambda_\Delta \int V_\Delta$, does not spoil conformal invariance [50] [51]. From the form of the correlation functions of section 4 and matrix models [37], [38] it follows that adding $\lambda_k \int T_k$ to the action for all $k \notin \frac{1}{\sqrt{2}}\mathbb{Z}$ and any λ_k doesn't spoil conformal invariance. We can calculate all correlators in a power series in λ_k (for fixed μ), and find sensible results. This seems to suggest that the exact non linear classical equations of motion of 2d string theory have the property that a tachyon field of the form $T(X) = \sum_{k \notin \frac{1}{\sqrt{2}}\mathbb{Z}} T_k$ (and trivial metric and other fields) is a solution. Of course, this can only be true up to field redefinitions, but even then it is

quite remarkable (a special case of this is the claim that Liouville (1.2) is a CFT). When we turn on an expectation value of a tachyon with one of the discrete momenta, the metric and other fields back react. The details of this back reaction haven't been worked out yet.

The higher genus correlation functions are another issue, which is still not resolved in the continuum formalism. In the matrix model approach, all order correlation functions were obtained in [37]. This is especially interesting in the light of the simple results obtained; as an example, the two point function was found to be given by:

$$\frac{\partial}{\partial \mu} \langle T_k T_{-k} \rangle = \frac{\Gamma(-\sqrt{2}|k|)}{\Gamma(\sqrt{2}|k|)} \text{Im} \left\{ e^{\frac{i\pi}{\sqrt{2}}|k|} \left[\frac{\Gamma(\frac{1}{2} + \sqrt{2}|k| - i\mu)}{\Gamma(\frac{1}{2} - i\mu)} - \frac{\Gamma(\frac{1}{2} - i\mu)}{\Gamma(-\sqrt{2}|k| + \frac{1}{2} - i\mu)} \right] \right\} \quad (5.1)$$

Higher point functions appear in [37]. To extract the result for given genus one should expand (5.1) in powers of $\frac{1}{\mu^2}$. It is challenging to derive the higher genus correlation functions in the continuum formalism, and even more challenging to understand the origin of the “non-perturbative” results, such as (5.1).

Acknowledgements

I would like to thank N. Seiberg for numerous discussions of the issues presented here; part of the work described in this review was done in collaboration with him, and the rest benefitted greatly from his suggestions and criticism. I would also like to thank M. Bershadsky and P. Di Francesco for collaboration and discussions. I have benefitted from discussions with T. Banks, R. Dijkgraaf, J. Distler, P. Freund, D. Gross, I. Klebanov, E. Martinec, G. Moore, N. Seiberg, S. Shatashvili, S. Shenker and H. Verlinde. I am grateful to the organizers of the spring school and ICTP for their hospitality. This work was partially supported by DOE grant DE-AC02-76ER-03072.

References

- [1] M. Green, J. Schwarz and E. Witten, *Superstring theory*, Cambridge Univ. Press (1987) and references therein.
- [2] A. Polyakov, *Gauge Fields and Strings*, Harwood Academic Publishers (1987).
- [3] E. Brezin, C. Itzykson, G. Parisi and J.B. Zuber, *Comm. Math. Phys.* **59** (1978) 35 ; D. Bessis, *Comm. Math. Phys.* **69** (1979) 147
- [4] V. Kazakov, *Phys. Lett.* **150B** (1985) 282; F. David, *Nucl. Phys.* **B257** (1985) 45; J. Ambjorn, B. Durhuus and J. Frohlich, *Nucl. Phys.* **B257** (1985) 433; V. Kazakov, I. Kostov and A. Migdal, *Phys. Lett.* **157B** (1985) 295.
- [5] D. Gross and A. Migdal, *Phys. Rev. Lett.* **64** (1990) 127; M. Douglas and S. Shenker, *Nucl. Phys.* **B335** (1990) 635; E. Brezin and V. Kazakov, *Phys. Lett.* **B236** (1990) 144.
- [6] E. Brezin, V. Kazakov, and Al. Zamolodchikov, *Nucl. Phys.* **B338** (1990) 673; D. Gross and N. Miljković, *Phys. Lett* **238B** (1990) 217; P. Ginsparg and J. Zinn-Justin, *Phys. Lett* **240B** (1990) 333; G. Parisi, *Phys. Lett.* **238B** (1990) 209.
- [7] T. Banks, M. Douglas, N. Seiberg and S. Shenker, *Phys. Lett.* **238B** (1990) 279.
- [8] N. Seiberg, Rutgers preprint RU-90-29 (1990).
- [9] G. Moore, N. Seiberg and M. Staudacher, Rutgers preprint RU-91-11 (1991).
- [10] G. Moore and N. Seiberg, Rutgers preprint RU-91-29 (1991).
- [11] A. Polyakov, *Phys. Lett.* **B103** (1981) 207, 211.
- [12] V. Knizhnik, A. Polyakov and A. Zamolodchikov, *Mod. Phys. Lett* **A3** (1988) 819.
- [13] F. David, *Mod. Phys. Lett.* **A3** (1988) 1651; J. Distler and H. Kawai, *Nucl. Phys.* **B321** (1989) 509.
- [14] S. Das, S. Naik and S. Wadia, *Mod. Phys. Lett.* **A4** (1989) 1033; J. Polchinski, *Nucl. Phys.* **B324** (1989) 123; T. Banks and J. Lykken, *Nucl. Phys.* **B331** (1990) 173.
- [15] A. Belavin, A. Polyakov and A. Zamolodchikov, *Nucl. Phys.* **B241**, 333 (1984).
- [16] T. Curtright and C. Thorn, *Phys. Rev. Lett.* **48** (1982) 1309; E. Braaten, T. Curtright and C. Thorn, *Phys. Lett.* **118B** (1982) 115; *Ann. Phys.* **147** (1983) 365.
- [17] J.L Gervais and A. Neveu, *Nucl. Phys.* **B209** (1982) 125; **B224** (1983) 329; **B238** (1984) 125.
- [18] D. Kutasov and N. Seiberg, *Nucl. Phys.* **B358** (1991) 600.
- [19] D. Kutasov and N. Seiberg, *Phys. Lett.* **B251** (1990) 67.
- [20] J. Polchinski, Texas preprints UTTG-19-90; UTTG-39-90 (1990).
- [21] A. Gupta, S. Trivedi and M. Wise, *Nucl. Phys.* **B340** (1990) 475.
- [22] A. Polyakov, *Mod. Phys. Lett.* **A6** (1991) 635.
- [23] A. B. Zamolodchikov *JETP Lett.* **43** (1986) 730; J. Cardy and A. Ludwig, *Nucl. Phys.* **B285[FS19]** (1987) 687; A. Cappelli, D. Friedan and J. Latorre, Rutgers preprint RU-90-43 (1990).

- [24] E. Witten, IAS preprint IASSNS-HEP-91/51 (1991).
- [25] E. Witten, Phys. Rev. **D44** (1991) 314.
- [26] J. Polchinski, Comm. Math. Phys. **104** (1986) 37; B. Maclain and B. Roth, Comm. Math. Phys. **111** (1987) 539; K. O' Brien and C. Tan, Phys. Rev. **D36** (1987) 1184.
- [27] D. Friedan, A. Kent, S. Shenker and E. Witten, unpublished; T. Banks, L. Dixon, D. Friedan and E. Martinec, Nucl. Phys. **B299** (1988) 613.
- [28] D. Friedan, E. Martinec and S. Shenker, Nucl. Phys. **B271** (1986) 93.
- [29] D. Gepner, Nucl. Phys. **B296** (1988) 757; PUPT-1121 (1989).
- [30] L. Dixon, Princeton preprint PUPT-1074 (1987).
- [31] A.B. Zamolodchikov, Phys. Lett. **117B** (1982) 87.
- [32] E. Martinec, Phys. Lett. **171B** (1986) 189.
- [33] D. Kutasov, G. Moore and N. Seiberg, unpublished.
- [34] E. Witten, Nucl. Phys. **B340** (1990) 281; J. Distler, Nucl. Phys. **B342** (1990) 523.
- [35] P. Di Francesco and D. Kutasov, Princeton preprint PUPT-1276 (1991).
- [36] M. Bershadsky and I. Klebanov, Phys. Rev. Lett. **65** (1990) 3088; N. Sakai and Y. Tanii, Tokyo preprint TIT/HEP-160 (1990).
- [37] G. Moore, Rutgers preprint RU-91-12 (1991).
- [38] D. Gross and I. Klebanov, Princeton preprint PUPT-1241 (1991); G. Mandal, A. Sengupta, and S. Wadia, IAS preprint, IASSNS-HEP/91/8 (1991); K. Demeterfi, A. Jevicki, and J.P. Rodrigues, Brown preprints BROWN-HET-795, 803 (1991); J. Polchinski, Texas preprint UTTG-06-91 (1991).
- [39] P. Di Francesco and D. Kutasov, Phys. Lett. **261B** (1991) 385.
- [40] N. Seiberg, unpublished.
- [41] B. Lian and G. Zuckerman, Yale preprint YCTP-P18-91 (1991); S. Mukherji, S. Mukhi and A. Sen, Tata preprint TIFR/TH/91-25 (1991); P. Bouwknegt, J. McCarthy and K. Pilch, preprint CERN-TH.6162/91 (1991).
- [42] J. Goldstone, unpublished; V. Kac, in *Group Theoretical Methods in Physics*, Lecture Notes in Physics, vol. 94 (1979); G. Segal, Comm. Math. Phys. **80** (1981) 301.
- [43] H. Sonoda and B. Zwiebach, Nucl. Phys. **B331** (1990) 592; **B336** (1990) 185.
- [44] S. Das and A. Jevicki, Mod. Phys. Lett. **A5** (1990) 1639.
- [45] A. Sengupta and S. Wadia, Tata preprint TIFR/TH/90-13 (1990).
- [46] D. Gross and I. Klebanov, Nucl. Phys. **B352** (1991) 671.
- [47] H. Ooguri and C. Vafa, Mod. Phys. Lett. **A5** (1990) 1389; Nucl. Phys. **B361** (1991) 469.
- [48] M. Bershadsky and D. Kutasov, Princeton preprint PUPT-1283 (1991).
- [49] M. Bershadsky and D. Kutasov, Phys. Lett. **266B** (1991) 345.
- [50] C. Callan, D. Friedan, E. Martinec and M. Perry, Nucl. Phys. **B262** (1985) 593.
- [51] T. Banks and E. Martinec, Nucl. Phys. **B294** (1987) 733.

TOPICS IN LIOUVILLE THEORY

L. ALVAREZ-GAUMÉ

*Theory Division, CERN
CH-1211 Geneva 23, Switzerland*

C. GÓMEZ^{* †}

*Département de Physique Théorique
Université de Genève
CH-1211 Genève 4, Switzerland*

^{*} Permanent address Instituto de Física Fundamental, CSIC. Serrano 123. Madrid 28006, Spain.

[†] Partially supported by the Swiss National Foundation.

1. INTRODUCTION

Liouville theory has a long history, especially in its classical properties. It is deeply related to the *Uniformization* problem. Poincaré, Klein, Koebe, Picard and many others studied the solutions to the classical Liouville equations in order to uniformize different types of surfaces. There are still many open problems in the classical theory of Fuchsian functions, and perhaps new methods in dealing with the quantization of Liouville theory will shed light on the classical problems. There is an abundant literature in Liouville theory (for a clear review and references see [1]), and in this lecture we will try to minimize the overlap as much as possible. We will start with general remarks to motivate the study of Liouville theory, later we will present the classical theory from a point of view not treated by previous lecturers in the school, and along the line of uniformization problems. The second lecture will study some of the exact results which can be obtained at the moment, in particular the scaling dimensions obtained by Knizhnik, Polyakov and Zamolodchikov [2] and the recent analysis by Bershadsky and Klebanov [3] of the spectrum of a minimal conformal model [4,5] coupled to gravity. In the third lecture we will study some of the results obtained in the computation of correlation functions in Liouville theory coupled to a minimal theory. We will not study the case of $c = 1$ because that has been the subject of previous lectures in the school.

When dealing with the theory of *random surfaces* one is often led to the quantum theory of the Liouville field. If we neglect the moduli for the time being, the Liouville mode represents the metric on the surface in the conformal gauge. In two dimensions it is known that many critical systems can be described in terms of random walks with some constraints. For example, polymers can be represented in terms of self-avoiding random walks and $c = 0, -2$ conformal field theories. The two dimensional Ising models and many IRF models can again be formulated in terms of special random walks. In three dimensional statistical systems one often encounters instead the notion of random surfaces replacing that of a random walk. To give an example in detail we may consider the three dimensional Ising model on a square three dimensional lattice. We will first present its dual theory, a Z_2 lattice gauge theory, and later we will show its equivalence with the 3D-Ising model. The partition function of a Z_2 lattice gauge theory is

$$Z = 2^{-3N} \sum_{\sigma_l} e^{\beta \sum_p \sigma(\partial p)} \quad (1.1)$$

to each link we associate a Z_2 variable $l \mapsto \sigma_l$, and for each plaquette we construct the plaquette variable $\prod_{l \in \partial p} \sigma_l$. N is the number of sites on the lattice. Expanding the

exponent we obtain

$$Z = 2^{-3N} \sum_{\sigma_l} \prod_p \cosh \beta (1 + \sigma(\partial p) \tanh \beta) \quad (1.2)$$

Using that $\sum_{\sigma_l} \sigma_l^n = 2$ for n even, and 0 for n odd, we find

$$Z = 2^{-3N} (\cosh \beta)^{3N} \sum_{\text{closed surfaces}} (\tanh \beta)^n \quad (1.3)$$

where n is the area of the closed surface; the number of plaquettes used in making it. From the expansion in (1.2) it is clear that each plaquette appears at most once in each surface, and that no more than four plaquettes can share the same link. The sum (1.3) can be written in a different way. A closed surface is characterized by the volumes it separates. If the original lattice is called L , let the dual lattice be L^* . To each dual lattice site we associate a Z_2 variable s_i , and the value $+1$ is assigned to it if it is external to a closed surface in the direct lattice, and -1 if it belongs to the inner volume. Each assignment of the s_i variables defines a closed surface on L up to a global $\pm$ sign. The area n can be written as

$$n = \frac{1}{2} \sum_{\langle i,j \rangle} (1 - s_i s_j) \quad (1.4)$$

The partition function (1.3) becomes now

$$\begin{aligned} Z &= \frac{1}{2} (\cosh \beta)^{3N} \sum_{\{s_i\}} (\tanh \beta)^{\frac{1}{2} \sum_{\langle i,j \rangle} (1 - s_i s_j)} \\ &= \frac{1}{2} (\cosh \beta)^{3N} (\tanh \beta)^{3N/2} \sum_{s_i} e^{-\frac{1}{2} \log \tanh \beta \sum_{\langle i,j \rangle} s_i s_j} \end{aligned} \quad (1.5)$$

If we define the new inverse temperature β^* by

$$\beta^* = -\frac{1}{2} \log \tanh \beta$$

or equivalently

$$\sinh 2\beta \sinh 2\beta^* = 1 \quad (1.6)$$

then the free energies of the two models are related according to

$$F(\beta) = F_{\text{Ising}}(\beta^*) + \frac{3}{2} \log \sinh 2\beta - \frac{1}{2} \log 2 \quad (1.7)$$

The above arguments show that a familiar model like the Ising model in three dimensions can be turned into a problem involving the sum over random surfaces. It is possible to give more examples where sums over such surfaces are relevant. In particular, the

theme motivating this school is the study of non-critical strings and two-dimensional gravity. In the continuum approach, the sum over surfaces is turned into a sum over topologies and over all possible geometries compatible with a given topology. When we use the conformal gauge to fix the diffeomorphism invariance, the metric dynamics is given in terms of the Liouville field theory. Only for critical strings the Liouville mode decouples [6]. If we want to study string theory in dimensions $d < 26$ we have to understand the quantum properties of this field theory. There are several regimes of interest. The classical limit is obtained as $d \mapsto -\infty$. The weak coupling limit occurs for $d \leq 1$ or $d \geq 25$. The strong coupling regime appears when $1 < d < 25$. This region remains largely unexplored. It is however clear that Liouville theory in the strong coupling region is related to many interesting problems in Statistical Mechanics and Quantum Field Theory apart from the field of Two-Dimensional Gravity, and any development in this direction in any of these subjects is likely to influence the others profoundly.

2. CLASSICAL LIOUVILLE THEORY AND UNIFORMIZATION PROBLEMS

We present the classical Liouville theory from the point of view of its original motivation back in the 19th century. Recent results by Zograf and Takhtadjan show that modern methods in field theory may shed new light in these old problems. In particular any method of computation of Liouville correlation functions may help in the explicit construction of uniformizing parameters for Riemann surfaces. We therefore begin with a problem apparently unrelated to Liouville theory.

To make the presentation as simple as possible we treat the Riemann sphere $S^2 = C \cup \{\infty\}$ minus n points $\{P_1, \dots, P_n\}$, $\Omega = S^2 \setminus \{P_1, \dots, P_n\}$ with $n \geq 3$. Otherwise we have the disk ($n = 1$) or the cylinder ($n = 2$) which are trivial cases as will become clear later in the lecture. The problem we want to analyze is the uniformization of Ω . The universal covering of this surface is the upper-half plane $\mathcal{H}$ [7]. Therefore we can find a mapping

$$\lambda : \mathcal{H} \rightarrow \Omega \quad z = \lambda(\tau) \quad (2.1)$$

where z is the standard coordinate on the sphere and τ is the coordinate on $\mathcal{H}$. The inverse mapping

$$\lambda^{-1} : \Omega \rightarrow \mathcal{H} \quad (2.2)$$

is holomorphic and locally single-valued, and can be analytically continued along every

path in Ω . In $\mathcal{H}$ the canonical metric is the Poincaré metric

$$ds^2 = \frac{d\tau d\bar{\tau}}{(\operatorname{Im}\tau)^2} \quad (2.3)$$

invariant under $SL(2, R)$ transformation

$$\tau \mapsto \frac{a\tau + b}{c\tau + d} \quad (2.4)$$

It is clear that the uniformization map is defined up to an $SL(2, R)$ transformation. When we use the inverse map (2.2) to lift a loop γ in Ω to a loop in $\mathcal{H}$ we will obtain a closed loop in $\mathcal{H}$ if the loop was in the trivial homotopy class in Ω . If, on the other hand, γ has winding number 1 around the puncture P_k it will lift to an open path $\tau \rightarrow T_k\tau$, $T_k \in SL(2, R)$. The transformation T_k or, rather, its conjugacy class, depends only on the homotopy class of the path γ . For punctures, the case of interest here, T_k is a parabolic transformation. Mathematically this fact determines the nature of P_k as a puncture. If the transformation T_k leaves the point α_k fixed ($\alpha_k \neq \infty$), it can be transformed to the form

$$(T_k\tau - \alpha_k)^{-1} = (\tau - \alpha_k)^{-1} + c_k \quad (2.5)$$

and if $\alpha_k = \infty$ then

$$T_k\tau = \tau - c_k \quad (2.6)$$

The set of generators $T_1, T_2, \dots, T_n$ obtained for all punctures generate a Fuchsian group satisfying a single relation $T_1 T_2 \dots T_n = 1$. It is known from elementary complex analysis that the Schwarzian derivative $\{\tau, z\}$ is invariant under $SL(2, R)$ transformations in the τ variable. The Schwarzian derivative of a function ψ with respect to z is defined as

$$\{\psi, z\} = \frac{\psi'''}{\psi'} - \frac{3}{2} \left(\frac{\psi''}{\psi'} \right)^2 \quad (2.7)$$

We can now quote the first result on uniformization (for details and references see [8]):

Theorem. If $z = \lambda(\tau)$ is the uniformization map $\lambda : \mathcal{H} \rightarrow \Omega$, then the Schwarzian

derivative $\{\tau, z\}$ is a meromorphic function of z with a rather simple form

$$\{\tau, z\} = \frac{1}{2} \sum_{k=1}^n \frac{1}{(z - p_k)^2} + \sum_{k=1}^n \frac{m_k}{z - p_k} \quad (2.8)$$

where the coefficients m_k , known as the auxiliary parameters, satisfy

$$\begin{aligned} \sum_k m_k &= 0 \\ \sum_k (2m_k p_k + 1) &= 0 \\ \sum_k (m_k p_k^2 + p_k) &= 0 \end{aligned} \quad (2.9)$$

The proof is carried out by first describing the transformation T_k as in (2.5) and noticing that near the puncture the function

$$\zeta(z) = e^{2\pi i c_k^{-1}(\tau - \alpha_k)^{-1}}$$

is single valued. Hence $\zeta(z) = a_1(z - p_k) + a_2(z - p_k)^2 + \dots$. Computing the Schwarzian derivative one arrives at (2.8) and (2.9) if one takes into account that ∞ is not chosen as a puncture, thus implying that $\{\tau, z\} \sim z^{-4}$ for large values of $|z|$. This asymptotic behavior together with the transformation properties of the Schwarzian derivative under coordinates redefinitions

$$\{\psi, \eta\} d\eta^2 = \{\psi, z\} dz^2 + \{z, \eta\} d\eta^2$$

strongly indicate that the Schwarzian derivative of the uniformization map can be thought of as the energy-momentum tensor of a classical field theory.

The Schwarzian $\{\tau, z\}$ can be written more conveniently if we express m_k as

$$m_k = \frac{1}{q_k - p_k}$$

then

$$\{\tau, z\} = \frac{1}{2} \sum_{k=1}^n \frac{1}{(z - p_k)^2} + \sum_{k=1}^n \frac{1}{(q_k - p_k)(z - p_k)} \quad (2.10)$$

Under Möbius transformations in Ω , $\{\tau, z\}$ transforms covariantly, if together with z and the punctures P_k we also transform the q_k 's as points on the plane.

Note that in (2.10) we can modify $\{\tau, z\}$ by adding

$$\eta(z) = \sum_k \frac{s_k}{z - p_k}$$

with the coefficients s_k satisfying

$$\sum_k s_k = 0 \quad \sum_k p_k s_k = 0 \quad \sum_k p_k^2 s_k = 0$$

under Möbius transformations one can choose the transformation properties of the variables s_k so that $\eta(z)$ transforms as a quadratic differential. This once more enhances our suspicion that the Schwarzian derivative is the energy-momentum tensor of a classical field theory. If we knew the uniformization map $z = \lambda(\tau)$ or its inverse, we could immediately calculate the constant negative curvature metric in Ω . It is given by

$$ds^2 = \rho^2 |dz|^2 \quad \rho(z) = (|\lambda'(\tau)| \operatorname{Im} \tau)^{-1} \quad (2.11)$$

$\rho(z)$ is single-valued and independent of the covering map. We can characterize ρ in a different, although equivalent way. The conformal factor $\rho(z)$ satisfies the following conditions:

- 1). It obeys the Classical Liouville equations of motion

$$\Delta \log \rho = \rho^2 \quad (2.12)$$

- 2). $\rho(z)$ is single-valued around the punctures.

- 3). Near a puncture p its behavior is

$$\rho(z) = \left[|z - p| \log \frac{1}{R|z - p|} \right]^{-1} + \text{const.} \quad (2.13)$$

- 4). The function $\rho(z)^{-1}$ satisfies a Fuchsian differential equation

$$\left(\frac{d^2}{dz^2} + \frac{1}{2} \{\tau, z\} \right) \rho^{-1} =$$

$$\left(\frac{d^2}{dz^2} + \frac{1}{4} \sum_{k=1}^n \frac{1}{(z - p_k)^2} + \frac{1}{2} \sum_{k=1}^n \frac{m_k}{z - p_k} \right) \rho^{-1} = 0 \quad (2.14)$$

Poincaré showed that this equation with the boundary conditions 3) has a unique solution. It is important to point out that we can find a set of uniformization parameters

m_k for which the equation (2.14) has as monodromy group a Fuchsian group but not necessarily the one uniformizing the Riemann sphere with n punctures at the points P_k . Unfortunately it is very difficult in general to obtain the solution of (2.14) even in terms of contour integrals. In terms of the standard Liouville field $\rho = e^{-\phi/2}$ (2.14) looks very close to the decoupling condition for a null vector at level two in Conformal Field Theory (CFT). In fact, the standard decoupling equations in CFT give rise in the classical limit to Fuchsian equations similar to (2.14). The only difference is that the parameters m_k in this case do not satisfy the uniformization constraints (2.9). We will see in a moment the physical reason for this. This strongly suggests that perhaps a suitable generalization of the Coulomb gas description of the minimal conformal models [9] should lead to useful integral representations of the uniformization map or of its inverse. We remind the reader that in the theory of ordinary differential equation the function ψ is the ratio of the two independent solutions $\eta_1\eta_2^{-1}$ where

$$\frac{d^2\eta}{dz^2} + \frac{1}{2}\{\tau, z\}\eta = 0 \quad (2.15)$$

This result is important in deriving condition 4). Since 1) and 4) are equivalent, we see that indeed the classical Liouville equation of motion is equivalent to the uniformization problem. The determination of the auxiliary parameters m_k is a very difficult and in general unsolved problem. In the case of three punctures (which for convenience can be taken to be at $\pm 1/2, \infty$) the solution of (2.14) is given in terms of the Legendre functions of order $-1/2$ and there are no auxiliary parameters. If the punctures are chosen according to $p_k = p^k$ where p is a primitive n -th root of unity, then it is not difficult to prove that $m_k = -p^{-k}/2$. The auxiliary parameters also play a crucial role in the description of the moduli space of spheres with punctures as has been shown recently by Zograf and Takhtadjan [10]. The basic idea is as follows: Consider the Liouville action

$$S = \int_{\Omega} \left(\frac{1}{2}(\partial\psi)^2 + e^{\psi} \right) - \text{regularization terms} \quad (2.16)$$

the regularization terms are needed because we will consider Liouville configurations satisfying the boundary conditions (2.13) at the punctures. Now if we solve the Liouville equation of motion with (2.13) boundary conditions and compute the action (2.16) , the auxiliary parameters are given by the variation of the classical action when the punctures change:

$$m_k = \frac{1}{2\pi} \frac{\partial S}{\partial p_i} \quad (2.17)$$

This result clarifies why the standard BPZ equations [4] are not uniformization equa-

tions. The parameters m_k appearing in the conformal field theory case satisfy in fact equation (2.17), but with S the free Coulomb gas action for static charges. Furthermore, the Weil-Petersson measure in moduli space is the second variation of the action

$$\left\langle \frac{\partial}{\partial p_i}, \frac{\partial}{\partial \bar{p}_j} \right\rangle = -\frac{\partial^2 S}{\partial p_i \partial \bar{p}_j} \quad (2.18)$$

With the Weil-Petersson form we can associate a Poisson bracket $\{, \}_{WP}$ with respect to which the moduli parameters p_i and the auxiliary parameters are canonically conjugate variables

$$\{m_i, m_j\}_{WP} = 0 \quad \{m_i, p_j\}_{WP} = \frac{i}{\pi} \delta_{ij} \quad (2.19)$$

This theorem of Zograf and Takhtadjan can be extended to the case of higher genus surfaces. It relates in a rather surprising way the geometry and topology of the moduli space of Riemann surfaces with the classical Liouville theory. The implications and consequences of this theorem are still largely unexplored. To summarize, we have explored the relations schematically expressed by the following arrows: Uniformization problem $\rightarrow$ Liouville equations of motion $\rightarrow S(p_1, p_2, \dots, p_n) \rightarrow$ accessory parameters $\rightarrow$ Weil-Petersson metric. We conclude this lecture with some remarks:

1). In this lecture we have talked exclusively about Liouville theory and Uniformization of Riemann surfaces. This led us to the energy-momentum tensor and to Fuchsian equations of the second order. One would expect a similar treatment to be related to the uniformization of holomorphic vector bundles on Riemann surfaces. This case should lead directly to W -algebras. One would expect that the analogue of the Zograf-Takhtadjan theorem should be obtained in terms of a Toda field theory and higher order Fuchsian equations. It is left as a challenge to the audience to prove this generalization.

2). The second order differential equation (2.14) is suspiciously close to a level two decoupling condition in degenerate conformal theories. It seems reasonable to expect a Coulomb gas treatment of the solutions to this equation. To reproduce the doubly logarithmic behavior however, we should presumably include the puncture operator $\phi e^{Q\phi/2}$ (see below) in some of the punctures.

3). One can derive the same results we have presented starting from an action as you have seen from previous lecturers. Choosing a fiducial metric $\hat{g}_{ab}$ on a given topological surface, we can write the physical metric as $g_{ab} = e^{\gamma\phi} \hat{g}_{ab}$. The Liouville

action is given by

$$S_I = \frac{1}{4\pi} \int \sqrt{\hat{g}} \left(\frac{1}{2} \hat{g}^{ab} \partial_a \phi \partial_b \phi + \frac{1}{\gamma} \phi R(\hat{g}) + \frac{\mu}{2\gamma^2} e^{\gamma\phi} \right) \quad (2.20)$$

Classically we can scale γ out of the action: $\phi \rightarrow \tilde{\phi}/\gamma$ leads to $S_L(\phi, \gamma) = \frac{1}{\gamma^2} S_L(\tilde{\phi}, \gamma = 1)$. Planck's constant is represented by γ^2 , and the classical limit is obtained by letting $\gamma \rightarrow 0$. The classical equation of motion for (2.20) is

$$\nabla_\mu \nabla^\mu = R(\hat{g}) + \frac{\mu}{2} e^{\gamma\phi} \quad (2.21)$$

This equation can be read as the conformal transformation interpolating between the fiducial metric and the constant curvature metric. Locally in isothermal coordinates we can always write the metric as $g_{ab} = e^{\gamma\phi} \delta_{ab}$. Using the equations of motion, the energy-momentum tensor takes the form

$$T(z) = -\frac{1}{2} \partial_z \phi \partial_z \phi + \frac{1}{\gamma} \partial_z^2 \phi \quad (2.22)$$

In isothermal coordinates, under a change of coordinates the field $\phi \rightarrow \phi'(z') = \phi(z) - \frac{1}{\gamma} \log |\partial f|^2$ and the energy-momentum tensor transforms according to

$$T(\phi) = T(\phi') \left(\frac{df}{dz} \right)^2 + \frac{1}{\gamma^2} \{f, z\} \quad (2.23)$$

Classically the central charge of the Virasoro algebra is $c = 12/\gamma^2$. Furthermore

$$T\left(\frac{1}{\gamma} \log \partial f\right) = \frac{1}{\gamma^2} \{f, z\} \quad (2.24)$$

The classical conformal dimension of a vertex operator $e^{\alpha\phi}$ is α/γ . From our uniformization arguments we can solve the Liouville equation of motion

$$e^{\gamma\phi} = \frac{16}{\mu} \frac{A'(z)B'(\bar{z})}{(1-AB)^2} \quad (B = \bar{A}) \quad (2.25)$$

Hence,

$$T(z) = \frac{1}{\gamma^2} \{A, z\}$$

The Schwarzian derivative is indeed the energy-momentum tensor of the Liouville theory. Another important property of Liouville theory is that it can be related to

a free field through a Backlund transformation [11]. The operator quantization of Liouville theory has also been carried out by J.L. Gervais and A. Neveu [12]. Working in Minkowski space and in light-cone coordinates, we introduce a new field ψ related to ϕ by

$$\begin{aligned}\partial_+\phi &= \partial_+\psi - \frac{\mu^{1/2}}{2\gamma} e^{\gamma(\phi+\psi)/2} \\ \partial_-\phi &= -\partial_-\psi + \frac{\mu^{1/2}}{2\gamma} e^{\gamma(\phi-\psi)/2}\end{aligned}\quad (2.26)$$

Then the equations

$$\partial_+\partial_-\phi = \frac{\mu}{4\gamma} e^{\gamma\phi} \quad \partial_+\partial_-\psi = 0$$

are equivalent. This transformation is in fact a canonical transformation generated by the generating functional

$$W[\phi, \psi] = \int_0^{2\pi} d\sigma \left(\phi\psi' - \frac{8\sqrt{\mu}}{\gamma^2} e^{\gamma\phi/2} \sinh \frac{\gamma\psi}{2} \right) \quad (2.27)$$

The Backlund transformation can be extended to the quantum theory to provide an operator solution to Liouville theory [11]. We should also mention to conclude that the field $A(z)$ appearing in (2.25) is not single-valued; it is defined up to a $SL(2, R)$ -transformation.. Using the explicit form (2.25) we can write some exponential field in a rather appealing form

$$\begin{aligned}e^{-j\gamma\phi} &= \left(\frac{16}{\mu}\right)^{-j} \left(\frac{1}{\sqrt{\partial A}} \frac{1}{\sqrt{\partial B}} - \frac{A}{\sqrt{\partial A}} \frac{B}{\sqrt{\partial B}} \right)^{2j} \\ &= \left(\frac{16}{\mu}\right)^{-j} \sum_{m=-j}^j \psi_m^j(z) \psi_m^j(\bar{z})\end{aligned}\quad (2.28)$$

where ψ_m^j is an $SL(2, R)$ representation of spin j . In particular for $j = 1/2$, $\psi_{\pm 1/2}$ satisfy a second order equation

$$\left(\frac{d^2}{dz^2} + \frac{\gamma^2}{2} T(z) \right) \psi_{\pm 1/2} = 0 \quad (2.29)$$

which is the starting point of another operator quantization of the Liouville theory (for details and references see [13]) closely related to quantum groups.

Next, we briefly review some of the problems appearing in the quantization of Liouville theory.

As lagrangian we use a modified version of (2.20)

$$\mathcal{L}_L = \frac{1}{8\pi}(\partial\phi)^2 + \frac{\mu}{8\pi\gamma^2}e^{\gamma\phi} + \frac{Q}{8\pi}\phi\sqrt{\hat{g}}R(\hat{g}) \quad (2.30)$$

with $Q = 2/\gamma + \gamma$. Were it not for the cosmological term, this would be a Feigin-Fuks lagrangian with central extension $c = 1 + 3Q^2$. To quantize (2.30) we may use standard canonical quantization. Denoting by $\phi(\sigma, \tau)$ and $\pi(\sigma, \tau)$ the co-ordinate and momentum fields, and using cylinder variables, we impose the equal-time commutator

$$[\phi(\sigma, \tau), \pi(\sigma', \tau)] = i\delta(\sigma - \sigma') \quad (2.31)$$

The Fourier decomposition at fixed time of the field variables is

$$\begin{aligned} \phi(\sigma, \tau) &= \phi_0(\tau) + \sum_{n \neq 0} \frac{i}{n} (a_n(\tau)e^{-in\sigma} + \tilde{a}_n(\tau)e^{in\sigma}) \\ \pi(\sigma, \tau) &= \pi_0(\tau) + \sum_{n \neq 0} \frac{i}{n} (a_n(\tau)e^{-in\sigma} + \tilde{a}_n(\tau)e^{in\sigma}) \end{aligned} \quad (2.32)$$

which are the ones we have used of a free field except that now we include an explicit time-dependence in the coefficients. This is due to the interactive nature of the Liouville lagrangian (2.30).

At fixed time, we may compute the energy-momentum tensor and from it the central extension. This is a conserved quantity, and in fact $c_L = 1 + 3Q^2$, the same value as if the cosmological constant was zero. Similarly, we may compute the conformal weight of the operators $e^{\alpha\phi(\sigma, \tau)}$, which is again the same as when $\mu = 0$:

$$\Delta_\alpha = \Delta(e^{\alpha\phi(\sigma, \tau)} :) = -\frac{1}{2}\alpha(\alpha - Q)$$

Up to this point, everything is very close to the standard Feigin-Fuks conformal field theories, but an important difference prevents us from defining in the usual way, *i.e. à la* BPZ, the quantum version of Liouville field theory. Recall that, in the Coulomb gas approach, we impose on correlators the condition of charge neutrality. We use the freedom of inserting screening charges in the correlator and thus we reproduce the Kac table. For the interacting theory, the situation is completely different: nothing prevents us from having non-vanishing correlators $\langle \prod e^{\alpha_i \phi(z_i, \bar{z}_i)} \rangle$ for arbitrary values of the charges α_i . In this sense, a BPZ-like description of the Liouville theory is hard to imagine.

Recall that the Fock space where we quantize the free field $X(z, \bar{z})$ decomposes into different irreducible representations of the Virasoro algebra, with highest weight determined by the free kinetic energy of the center of mass momentum. In the interacting case, the center of mass motion is not free anymore and, in order to know the value of its energy, we should solve the appropriate fully interacting Schrödinger problem. A way to simplify the computation is to use the so-called mini-superspace approximation, where we drop all dependence on σ . It is expected that the qualitative features of the solution survive this simplification. Here we will follow closely the clear presentation in [1].

In the mini-superspace approximation, then, the functional Schrödinger equation takes the following form:

$$\left(\frac{1}{2} \pi_0^2 + \frac{\mu}{8\gamma^2} e^{\gamma\phi_0} + \frac{Q^2}{8} \right) \Psi[\phi_0] = E\Psi[\phi_0] \quad (2.33)$$

From (2.33) we can guess the structure of the physical Hilbert space for the Liouville theory:

$$\mathcal{H} = \bigoplus_{p>0} \left[\mathcal{H}_{E(p)} \otimes \mathcal{H}_{E(p)} \right] \quad (2.34)$$

with $E(p) = p^2/2 + Q^2/8$ the energy eigenvalues of the Schrödinger equation. In the free theory, we may obtain the state $|p\rangle$, describing a string with center of mass momentum p , by acting on the vacuum state $|0\rangle$ with the operator $e^{ipX(z, \bar{z})} \Big|_{z=\bar{z}=0}$. The problem at hand is that we do not have the state $|0\rangle$ at our disposal. In other words, due to the effect of the interaction we have lost the connection between operators and states. All we have is, on one hand, states $|p\rangle$ (and their oscillator descendants, which we do not see in the mini-superspace approximation) with “energy” $E(p)$, and, on the other hand, operators $e^{\alpha\phi}$. The immediate questions we might ask is when are the operators $e^{\alpha\phi}$ well-defined on the Hilbert space (2.34) of the interacting theory, and how to compute their expectation values. The answer to these questions is, at the time of this writing, unknown. Here we shall present the elements for a possible solution using the functional integral representation.

As already mentioned, one of the main problems with quantum Liouville theory is that we do not have a well-defined vacuum state on which to act with the operators $e^{\alpha\phi}$ (or others) in order to obtain highest weight representations of the Virasoro algebra. A way out is, of course, to use the operator formalism to define the vacuum as the state

associated, through the lagrangian (2.30), with the once-punctured Riemann sphere:

$$|0\rangle^{(L)} = |\Sigma_{0,1}\rangle$$

with the wave-function of $|\Sigma_{0,1}\rangle$ defined by

$$\Psi(f) = \int \mathcal{D}\phi \Big|_{\phi|_{\partial\mathbb{D}}=f} \exp - \int \mathcal{L}_L(\phi) \quad (2.35)$$

If we accept for the time being this definition of the vacuum, the result of acting with $e^{\alpha\phi}$ on it will be the state associated with the twice punctured Riemann sphere, where the operator $e^{\alpha\phi}$ is inserted at one of the punctures. The associated wave-functional is given by

$$\Psi^{(\alpha)}(f) = \int \mathcal{D}\phi \Big|_{\phi|_{\partial\mathbb{D}}=f} e^{\alpha\phi} \exp - \int \mathcal{L}_L(\phi) \quad (2.36)$$

This procedure has, nevertheless, some problems. First, of course, we should verify that the proposed wave-functionals do belong to the Hilbert space $\mathcal{H}$. The well-defined operators $e^{\alpha\phi}$ are those which map states in $\mathcal{H}$ to states in $\mathcal{H}$. A simple way to characterize the states in the Hilbert space $\mathcal{H}$ is to require the associated wave-functionals to be normalizable.

In order to address the issue of normalizability, it is convenient to consider generic correlators first. Let us try to evaluate a correlator of exponentials of the Liouville field by separating out the area $A = \int e^{\gamma\phi}$ from the functional integration:

$$\begin{aligned} \langle \prod_i e^{\alpha_i \phi(z_i, \bar{z}_i)} \rangle &= \int \mathcal{D}\phi \prod_i e^{\alpha_i \phi(z_i, \bar{z}_i)} \exp - \int \mathcal{L}_L \\ &= \int dA \langle \prod_i e^{\alpha_i \phi(z_i, \bar{z}_i)} \rangle_A e^{-\frac{\mu}{8\pi\gamma^2} A} \end{aligned} \quad (2.37)$$

The correlator in the last line should be evaluated for fixed area. On the sphere, we obtain from (2.37) the following dependence on the area of the correlator:

$$\langle \prod_i e^{\alpha_i \phi(z_i, \bar{z}_i)} \rangle \sim \left[\int dA A^{-1+\nu/\gamma} e^{-\mu A/8\pi\gamma^2} \right] \langle \prod_i e^{\alpha_i \phi(z_i, \bar{z}_i)} \rangle_{A=1} \quad (2.38)$$

where we have set

$$\nu = \left(\sum_i \alpha_i - \frac{Q}{2} \right)$$

Clearly, the expression (2.38) is divergent for $\nu < 0$.

This analysis for the case of a single exponential, i.e. for the wave-functionals above, implies that they are normalizable only if $\nu > 0$. For instance, the “vacuum state” defined by (2.36) has $\nu < 0$ and therefore is not normalizable, and does not belong to the Hilbert space $\mathcal{H}$ in equation (2.34). Moreover, the wave-functionals $\Psi^{(\alpha)}(f)$ belong to $\mathcal{H}$ iff $\alpha > Q/2$, since only then are they normalizable. If $\alpha < Q/2$, then the wave-functional $\Psi^{(\alpha)}$ is not normalizable and the state (2.36) is not in the Hilbert space $\mathcal{H}$.

The different behaviour of the wave-functionals $\Psi^{(\alpha)}$ for α larger or smaller than $Q/2$ was introduced by Seiberg under the name of macroscopic and microscopic states, respectively. A simple but useful way to see the relation between operators $e^{\alpha\phi}$ with $\alpha > Q/2$ and states in $\mathcal{H}$ is to define a momentum p associated with α via the relation

$$\alpha = \frac{Q}{2} + ip \quad (2.39)$$

This implies then that the energy $E(p)$ of the state $|p\rangle \in \mathcal{H}$ is equal to the conformal weight Δ_α of the operator $e^{\alpha\phi}$. From the identification (2.39) it is clear that operators with $\alpha < Q/2$ carry imaginary momentum. For imaginary momenta, the Liouville theory becomes effectively equivalent to a $c < 1$ conformal field theory (all monodromies become elliptic).

In good Liouville logic, we should consider good operators only $e^{\alpha\phi}$ with $\alpha > Q/2$: these are the “macroscopic states”. Unfortunately, even these are not free of problems. A major one concerns locality. Indeed, the definition of locality (relative to the two-dimensional surface) in Liouville theory is a thorny issue. To give meaning to the local co-ordinates, we need to use the metric $g = e^\phi \hat{g}$ which contains, itself, the Liouville field. Thus if the operator $e^{\alpha\phi}$ is to be local, then the wave-functional $\Psi^{(\alpha)}$ should be, in the mini-superspace approximation, close to a δ -function localized around $\phi_0 = \infty$. This makes the wave-functional irredeemably not normalizable: it is the origin of the nomenclature “microscopic” for the states with $\alpha < Q/2$. We should use microscopic or macroscopic operators depending on what we wish to describe. For instance, if we want to use the Liouville field to define a string theory with $c_{\text{matter}} < 26$, then in the spectrum we should consider microscopic states. We shall discuss this point further in the last lecture.

A different approach to the quantization of the Liouville field is to use quantum group techniques, the main ideas of which we now sketch. Essentially, instead of (2.32) we shall use a different expansion (2.28) of the Liouville field.

In the quantum group approach, instead of imposing the Heisenberg quantization rule, we shall replace the internal symmetry group $SL(2)$ by its quantum deformation

$SL(2)_q$, for an appropriate value of q . The fields $\psi_m^j(z)$ will be required, thus, to transform as spin- j multiplets of $SL(2)_q$:

$$\psi_{m_1}^{j_1}(z_1) \psi_{m_2}^{j_2}(z_2) = R_{m_1, m_2; m'_1, m'_2}^{j_1, j_2} \psi_{m'_2}^{j_2}(z_2) \psi_{m'_1}^{j_1}(z_1)$$

with R^{j_1, j_2} the R -matrix of $SL(2)_q$ in the representation $j_1 \otimes j_2$.

The quantum operator $e^{-j\gamma\varphi}$ is defined by (2.28), with the understanding that the fields $\psi_m^j(z)$ and $\bar{\psi}^{jm}(\bar{z})$ transform as irreps of spin j and $\bar{j}$ of the quantum group. Already at the level of (2.28) we can see the difference between microscopic and macroscopic states. Actually, the condition for the operators $e^{\alpha\varphi}$ to be well-defined is stronger in this quantization: α should be equal to $-j\gamma$, with $2j \in \mathbf{N}$. If j is not a positive integer or half-integer, then the expansion of $(1 - AB)$ is still possible, but it contains an infinite number of terms:

$$\exp x\gamma\varphi = \left(\frac{16}{\mu}\right)^x \sum_{m=x}^{\infty} \psi_m^x(z) \bar{\psi}^{xm}(\bar{z}) \quad (2.40)$$

Attempts to play with infinite-dimensional representations of $SL(2)_q$ in order to give meaning to this expression have not met with success.

The correct interpretation of the quantum versions of (2.28) and (2.40) highlights the invariance under monodromy of the right-hand sides: the physical quantum operators are monodromy-invariant combinations of the chiral components. The chiral components, up to a possible zero-mode, define a typical BPZ chiral structure, namely a fusion algebra given by the decomposition rules of $SL(2)_q$:

$$j_1 \times j_2 = \begin{cases} \sum_{|j_1-j_2|}^{j_1+j_2} j & \text{if } q \text{ is not a root of unit} \\ \sum_{|j_1-j_2|}^{\min(j_1+j_2, p-2-j_1-j_2)} j & \text{if } q^p = 1, p \text{ a natural number larger than 3} \end{cases}$$

The fusion algebra of (2.40) on the other hand, is a very bad one from a physical point of view because it does not contain an identity (we set $x = \gamma j$, $2j$ a positive integer):

$$j_1 \times j_2 = \sum_{j_1+j_2}^{\infty} j$$

and hence all the two-point functions vanish. Moreover, there is no natural candidate for an energy-momentum tensor.

Summarizing, the quantum group approach yields a BPZ picture for the microscopic states, but it does not allow us yet to deal with the interesting Liouville part of the problem, *i.e.* the macroscopic operators.

3. SOME EXACT RESULTS IN LIOUVILLE THEORY

In this lecture we will study two-dimensional Conformal Field Theories coupled to two dimensional gravity. For these models, each primary field becomes gravitationally dressed by the Liouville field. If Ψ_0 is a primary field of the matter theory with conformal dimension h , the dressing takes the form

$$\Psi = \Psi_0 e^{\beta\phi} \quad (3.1)$$

and β is determined by requiring the combined conformal dimension of (3.1) to be equal to 1. In this way the field Ψ is a $(1, 1)$ form and can be integrated over the surface. If the conformal theory has central extension c , the value of Q is determined to be

$$Q = \sqrt{\frac{25 - c}{3}} \quad (3.2)$$

so that the total contribution to the central charge of the Virasoro algebra of the ghosts, Liouville and matter fields cancel. The classical limit is obtained when $c \rightarrow \infty$ and this determines γ as a function of c

$$\gamma = \frac{Q - \sqrt{Q^2 - 8}}{2} = \frac{\sqrt{25 - c} - \sqrt{1 - c}}{\sqrt{12}} \quad (3.3)$$

The dressing factor β in (3.1) is determined by solving the equation

$$h_0 - \frac{1}{2}\beta(\beta - Q) = 1$$

There are two solutions. Choosing the one agreeing with the classical limit, we obtain

$$\beta = \frac{Q}{2} - \sqrt{\frac{Q^2}{4} - 2 + 2h_0} = \frac{\sqrt{25 - c} - \sqrt{1 - d + 24h_0}}{\sqrt{12}} \quad (3.4)$$

The parameter β determines the scaling properties of the gravitationally dressed primary fields in the presence of gravity. To understand how critical exponents are defined,

let us recall the definition of the string susceptibility. The same argument extends to other cases as well. In the problem of two-dimensional quantum gravity we are interested in computing the partition function and its dependence on the cosmological constant μ which can be thought of as the chemical potential for the area. In terms of the area microcanonical ensemble the partition function is written as

$$Z(\mu) = \sum_A Z(A) e^{-\mu A} \quad (3.5)$$

The critical behavior of $Z(\mu)$ is determined by the large area behavior of $Z(A)$:

$$Z(A)_{A \rightarrow \infty} \sim e^{\mu_c A} A^{\Gamma-3} \quad (3.6)$$

the exponent Γ is by definition the string susceptibility, and μ_c is the critical value of the cosmological constant. Substituting (3.6) into (3.5) we obtain near μ_c

$$Z(\mu) \sim (\mu - \mu_c)^{2-\Gamma} \quad (3.7)$$

Similarly, for the two point function of a field Φ we can define its critical exponent by

$$e^{-\mu A} Z_{\Phi\Phi}(A)_{A \rightarrow \infty} \sim A^{2-2h_\Phi+\Gamma-3} e^{-(\mu-\mu_c)A} \quad (3.8)$$

hence for $\mu \sim \mu_c$

$$Z_{\Phi\Phi}(\mu) \sim (\mu - \mu_c)^{2h_\Phi-\Gamma}$$

For normalized one-point functions at fixed area $F_\Phi(A) \sim A^{1-h}$. These definitions can be found among many other things in the seminal paper [2]. The simplest result in Liouville theory in the continuum is the determination of the gravitationally dressed dimensions (3.8) for the minimal conformal models [4,5]. In the next subsection we give a detail account on some general results for these theories.

3.1 SCALING DIMENSIONS AND SPECTRUM

There are few exact results available in the continuum description of the coupling of a Conformal Field Theory (CFT) to $2D$ -gravity. Their derivation can be understood through a rather simple argument. In this paper we concentrate on the coupling of

a minimal conformal model [4] to Liouville theory. These models are defined by two coprime integers (p', p) ; $p' > p$ and have a central charge

$$c = 1 - 6 \frac{(p' - p)^2}{p'p} = 1 - 3\alpha_0^2 \quad (3.9)$$

The primary fields have dimensions

$$\Delta_{m',m} = \frac{1}{8}[(m\alpha_+ + m'\alpha_-)^2 - \alpha_0^2] = \frac{1}{4p'p}[(mp' - m'p)^2 - (p' - p)^2] \quad (3.10)$$

with

$$\begin{aligned} \alpha_+ + \alpha_- &= \alpha_0 & \alpha_+ \alpha_- &= -2 \\ \alpha_+ &= \sqrt{\frac{2p'}{p}} & \alpha_- &= -\sqrt{\frac{2p}{p'}} \end{aligned} \quad (3.11)$$

The range of m, m' is

$$1 \leq m \leq p-1 \quad 1 \leq m' \leq p'-1 \quad m'p - mp' \geq 0 \quad (3.12)$$

These models are very special because the Verma modules $V_{m',m}$ generated by $\Delta_{m',m}$ have an infinite number of null vectors. Defining

$$\begin{aligned} a_{m',m}(k) &\equiv \frac{1}{4p'p}[(2p'pk + mp' - m'p)^2 - (p' - p)^2] \\ b_{m',m}(k) &\equiv \frac{1}{4p'p}[(2p'pk + mp' + m'p)^2 - (p' - p)^2] \end{aligned} \quad (3.13)$$

the null vector structure of the module $V_{m',m}$ is described by the diagram [14,15]

$$\begin{array}{ccccccc} & & a(0) \rightarrow b(0) & \rightarrow a(1) & \rightarrow b(1) & \rightarrow a(2) & \rightarrow \dots \\ E_{m',m}(p',p) & & \searrow & \times & \times & \times & \times \\ & & b(-1) \rightarrow a(-1) & \rightarrow b(-2) & \rightarrow a(-2) & \rightarrow \dots \end{array} \quad (3.14)$$

The meaning of this graph is as follows. Each node represents a Verma module with $a(k)$ or $b(k)$ as the dimension of its highest weight vector, and an arrow connecting two spaces $E \rightarrow F$ means that the module F is contained in the module E . This

diagram allows us to compute the Virasoro character of the irreducible representation $L_{m',m}(p',p)$ obtained by eliminating the null vectors; the result is [15]

$$\begin{aligned} \chi_{m',m} &= \text{tr} q^{L_0 - c/24} = \\ &= \frac{1}{\eta(q)} \sum_{k \in \mathbb{Z}} \left(q^{\frac{1}{4p'p}(2p'pk + mp' - m'p)^2} - q^{\frac{1}{4p'p}(2p'pk + mp' + m'p)^2} \right) \end{aligned} \quad (3.15)$$

where $\eta(q)$ is the Dedekind η -function. All the properties of the minimal models we have described can be represented explicitly in terms of a Coulomb gas formulation [16].

In the evaluation of n -point functions of gravitationally dressed operators we can split the computation into the matter and the Liouville parts as long as we assume that there is no mixing between both sets of fields in the action. The Liouville part for a genus h surface (we follow the notation of [1]) is

$$\left\langle \prod_i e^{\beta_i \phi(P_i)} \right\rangle = \int d\mu(\tau) \mathcal{D}\phi e^{-\frac{1}{8\pi} \int (\nabla_i \phi \nabla^i \phi + QR\phi + \mu \Psi_{\min} e^{\beta_{\min} \phi})} \prod_i e^{\beta_i \phi(P_i)} \quad (3.16)$$

Here $d\mu(\tau)$ represents the measure over the moduli space of Riemann surfaces of genus h , and we lump there as well the contribution due to the ghost determinant. $\Psi_{\min}$ is the primary field in the CFT with lowest dimension, and $\beta_{\min}$ is its gravitational dressing [2]. For a unitary theory this field is the identity. To evaluate (3.16) we insert an "area constraint" δ -function and integrate over all areas. As we have seen to compute the string susceptibility and gravitational dimensions we need to introduce these δ -function constraints. Changing the order of integration we obtain

$$\begin{aligned} \left\langle \prod_i e^{\beta_i \phi(P_i)} \right\rangle &= \int d\mu(\tau) \mathcal{D}\phi \int_0^\infty dA e^{-\mu A/8\pi} \delta\left(\int \Psi_{\min} e^{\beta_{\min} \phi} - A\right) \\ &\quad e^{-\frac{1}{8\pi} \int (\nabla_i \phi \nabla^i \phi + QR\phi)} \prod_i e^{\beta_i \phi(P_i)} \end{aligned} \quad (3.17)$$

Next we eliminate the δ -function by splitting the integration over ϕ into the zero mode and the non-zero mode pieces, $\phi = \phi_0 + \tilde{\phi}$. Using the Gauss-Bonnet theorem and

standard properties of δ -functions yields

$$\langle \prod_i e^{\beta_i \phi(P_i)} \rangle = \frac{1}{\beta_{\min}} \int_0^\infty dA e^{-\mu A/8\pi} A^{\frac{1}{\beta_{\min}}(\sum \beta_i - Q(1-h)) - 1} \\ \int d\mu(\tau) \mathcal{D}\tilde{\phi} \tilde{A}^{-\frac{1}{\beta_{\min}}(\sum \beta_i - Q(1-h))} e^{-\frac{1}{8\pi} \int (\nabla_i \tilde{\phi} \nabla^i \tilde{\phi} + QR\tilde{\phi})} \prod_i e^{\beta_i \tilde{\phi}(P_i)} \quad (3.18)$$

where

$$\tilde{A} = \int \Psi_{\min} e^{\beta_{\min} \tilde{\phi}} \quad (3.19)$$

and

$$s = \frac{1}{\beta_{\min}} (Q(1-h) - \sum \beta_i) \quad (3.20)$$

We obtain finally

$$\langle \prod_i e^{\beta_i \phi(P_i)} \rangle = \frac{1}{\beta_{\min}} \left(\frac{\mu}{8\pi} \right)^s \Gamma(-s) \int d\mu(\tau) \mathcal{D}\tilde{\phi} \tilde{A}^s e^{-\frac{1}{8\pi} \int (\nabla_i \tilde{\phi} \nabla^i \tilde{\phi} + QR\tilde{\phi})} \prod_i e^{\beta_i \tilde{\phi}(P_i)} \quad (3.21)$$

When the theory is unitary, $\Psi_{\min} = 1$, the action in the integrand of (3.16) is the standard Liouville action, and μ is the cosmological constant. In order to facilitate the comparison with matrix model results it is convenient to use (3.16) instead of the ordinary Liouville action. However, in the computation of three-point functions we will work with the standard action. This is equivalent to fine tuning the coefficients of the most relevant operators (those of negative dimensions) in the non-unitary case. Using (3.21) we can read off the gravitational dimensions of fields on any genus surface and obtain the results of [2,18,17]. We see explicitly that the string susceptibility at genus zero is given by

$$\Gamma = 2 - \frac{Q}{\gamma}$$

and at arbitrary genus h

$$2 - \Gamma(h) = (2 - \Gamma)(1 - h) \quad (3.22)$$

The gravitational dimensions also follow from (3.21)

$$h = 1 - \frac{\beta}{\gamma} \quad h - h_0 = \frac{\gamma^2}{2} h(1 - h) \quad (3.23)$$

For minimal models it is easy to check that the β factor takes the form

$$\beta_{m',m} = \frac{Q}{2} - \frac{|mp' - m'p|}{\sqrt{2pp'}} \quad (3.24)$$

and in order to compare with matrix models

$$\frac{\beta}{\beta_{min}} = \frac{p + p' - |mp' - m'p|}{p + p' - 1} \quad (3.25)$$

Furthermore, as noted in [3] for genus 1 and $\beta_i = 0$, the remaining functional integral is a free field functional integral and it can be carried out easily. So far we have ignored the ghost contribution, but it should now be included. The result of ref. [3] is that the partition function is proportional to

$$\int_{\mathcal{F}} d^2\tau \tau_2^{-3/2} |\eta(q)|^2 Z_M(q, \bar{q})$$

$$Z_M(q, \bar{q}) = \frac{1}{|\eta(q)|^2} \sum_{\substack{1 \leq m \leq p-1 \\ 1 \leq m' \leq p'-1}} |\chi_{m',m}|^2 \quad (3.26)$$

where the matter characters appear in (3.15). Since the η factors cancel, there are minus signs in the expansion of the integrand which are interpreted as ghost number counting contributions $(-1)^g (-1)^{\bar{g}}$. Hence, it seems that the would-be null states in CFT resuscitate, although with ghost dressing. We can therefore distinguish the dressing of the standard primary fields which only get Liouville contributions from those of the new states which also get ghost dressing. The ghost structure of these new states is best understood in terms of the BRST analysis of Lian and Zuckerman [19].

3.2 BRST ANALYSIS

In the full theory of Liouville + matter and ghosts, the total value of the central charge is $c = 0$, and we can define a BRST charge Q by

$$Q = \oint \frac{dz}{2\pi i} J(z)$$

$$J(z) =: (T_L(z) + T_M(z) + \frac{1}{2} T_{gh}(z)) c(z) : \quad (3.27)$$

The energy-momentum tensor for Liouville is

$$T_L(z) = -\frac{1}{2}(\partial\phi)^2 + \frac{Q}{2}\partial^2\phi \quad c_L = 1 + 3Q^2 \quad (3.28)$$

and the conformal dimension of exponential operators is

$$\Delta(e^{\beta\phi}) = -\frac{1}{2}\beta(\beta - Q) \quad Q = \sqrt{\frac{25 - c_M}{3}} \quad (3.29)$$

We will label the Fock spaces generated by the Liouville oscillators out of the state created by the vertex operator $e^{\beta\phi}$ by $\mathcal{F}_{\beta,Q}$. The general analysis of the cohomology of the BRST charge can be carried out in terms of an arbitrary representation of the Virasoro algebra with energy-momentum tensor $T(z)$ and central charge $26 - c_M$. Let Λ^{bc} be the ghost Fock space, which can be thought of as a space of semi-infinite forms. If $L_{m',m}(p',p)$ is the irreducible Virasoro module associated to the primary field with dimension (2.2) and central charge (2.1), and V is a Virasoro module with central charge $26 - c_M$, the first result of Lian and Zuckerman [19] implies that the cohomology groups of Q (the BRST closed modulo exact states) acting on the space $\Lambda^{bc} \otimes V \otimes L_{m',m}(p',p)$, which will be denoted by $H_*^{\text{abs}}(V)$, are non-vanishing only when V is a vector space in the following sequence of embeddings

$$\begin{array}{ccccccc} & 1 - a(0) \leftarrow 1 - b(0) & \leftarrow 1 - a(1) & \leftarrow 1 - b(1) & \leftarrow 1 - a(2) & \leftarrow \dots \\ \tilde{E}_{m',m}(p',p) & \nwarrow & \times & \times & \times & \times \\ & 1 - b(-1) \leftarrow 1 - a(-1) & \leftarrow 1 - b(-2) & \leftarrow 1 - a(-2) & \leftarrow \dots \end{array} \quad (3.30)$$

where again every node in the graph represents a Verma module whose highest weight conformal dimension is given, and the arrows have exactly the same meaning as in (3.14). This exhibits a very interesting duality noticed in [14] between representations of Virasoro characterized by c, Δ and those characterized by $26 - c, 1 - \Delta$. The fact that the arrows in (3.30) are reversed with respect to those in (3.14) implies that the module $1 - a(0)$ has no null vectors, the first two joined with one arrow to $1 - a(0)$ have this module as a common null submodule, etc.

If we particularize this result to the case when V is the Fock module $\mathcal{F}_{\beta,Q}$, the result of [19] needed here is that the dimensions and ghost numbers of the absolute BRST cohomology groups are given by

$$\dim H_{n+1/2}^{\text{abs}}(\mathcal{F}_{\beta,Q}) = \delta_{n \pm d_\beta, 0} + \delta_{n \pm d_\beta, -1} \quad (3.31)$$

In this formula d_β is the distance along the arrows in (3.30) from the selected node in the graph to the tip $1 - a(0)$. The $+$ (resp. $-$) sign occurs when $\beta \leq Q/2$ (resp. $\beta > Q/2$).

As a final remark we should mention that in the comparison between the gravitational scaling dimensions that one finds in the matrix models and those obtained from (3.21) one is still missing in the continuum a detailed understanding of a set of operators whose gravitational scaling dimensions coincide with those one would obtain if we included the boundaries of the Kac table, in other words, states with dimensions $\Delta_{m',0}, m' \neq 0 \pmod{p}$. These states are also peculiar from the point of view of the Virasoro modules analyzed in [14]. They only have one generating null vector, and the embedding diagram for null vectors contains a single line of arrows. It should be interesting to find out whether one can construct modular invariants with their associated character, and if so, why such characters do not appear in the one-loop results of ref. [3].

4. THREE POINT FUNCTIONS IN THE CONTINUUM

In this section we would like to comment on the computation of three point functions. The procedures advocated so far imply some analytic continuation. The computations are rather involved, but the ideas are simple. We will present the general outline of the computation and give the result. The relevant references with the details are [20,21,22,23,24]. Recall one of the main results from section 3.1. The final result for the correlator of n -point functions was (3.21)

$$\langle \prod_i e^{\beta_i \phi(P_i)} \rangle = \frac{1}{\beta_{\min}} \left(\frac{\mu}{8\pi} \right)^s \Gamma(-s) \int d\mu(\tau) \mathcal{D}\tilde{\phi} \tilde{A}^s e^{-\frac{1}{8\pi} \int (\nabla_i \tilde{\phi} \nabla^i \tilde{\phi} + Q R \tilde{\phi})} \prod_i e^{\beta_i \tilde{\phi}(P_i)}$$

Previously we analyzed the case when $h = 1$ and all the β 's vanish. Suppose we want to consider the three-point function. This is what we need, for example, if we want to compute the fusion rules; or, rather, the selection rules for the three-point function in the fully dressed theory. If (3.20) were a positive integer, we could evaluate (3.21) using standard free field methods. When s is not an integer the computation is rather difficult and there are no methods available yet that we could apply. The proposal of Goulian and Li [20] was to analytically continue in s . Using the results of [16] one can compute the three-point function and then analytically continue to the general case. This is an analytic continuation in the central charge of the matter Virasoro algebra. Similar methods were employed later in [21,22]. In [23] the Coulomb gas techniques [16] are applied to the Liouville theory and an analytical continuation procedure is introduced to define the correlators with a negative number of screening charges. The results of all these references is that the three point function factorizes into leg factors.

Define the function

$$\Delta(x) = \frac{\Gamma(x)}{\Gamma(1-x)} \quad (4.1)$$

In the Coulomb gas representation every primary field is represented by a vertex operator $e^{i\alpha\psi}$, and its gravitational dressing is given by a Liouville vertex operator $e^{\beta\phi}$. We can construct a Minkowski two-vector $p = (\beta, \alpha)$ and we can define its square $p^2 = \beta^2 - \alpha^2$. Then the three point function of the fully dressed field is given by

$$\langle V_{p_1}(\infty)V_{p_2}(1)V_{p_3}(0) \rangle = \text{constant} \mu^s \prod_{a=1}^3 \Delta\left(\frac{1}{2}p^2\right) \quad (4.2)$$

It is believed that a similar structure will be found for the integrated n -point function, although a complete proof is not yet available. It is important to notice that the leg factors in (4.2) have poles when the matter fields live on the boundary of the Kac table, in other words, when the matter field has the labels $(0, m)$. These are analogous to the leg poles discussed in [25]. These states do not appear in the partition function computation in [3] or in the BRST analysis in [19]. One possible explanation is that these states are redundant and they should not appear in the continuum description of the minimal models coupled to gravity [26]. Taken literally, (4.2) seems to imply that there are no selection rules (fusion rules) in the three point correlators and that the one- and two-point functions do not satisfy reasonable properties (one-point functions should vanish and two-point functions should be diagonal). Another puzzling aspect is the apparent recoupling of would-be null vectors in a CFT, as seen from the computation of Bershadsky and Klebanov presented in the previous section. These puzzles can be resolved if we take into account the fact that we are trying to compute string correlation functions and, as shown by Lian and Zuckerman, the ghost number assignments are very different from those of the critical strings. In the next subsection we will show that if we consider correlation function of microscopic states on arbitrary Riemann surfaces involving some of the would-be null vectors, the answer one obtains is zero, and furthermore one can recover the standard fusion rules before the theory was coupled to gravity.

4.1 COMPUTATIONAL RULES

In this section, we wil present the explicit computation of some correlators at genus zero. The generic correlation functions we shall study are of the form

$$\langle \prod_i V_{m'_i m_i}^{(\pm)}(z_i, \bar{z}_i) \rangle \quad (4.3)$$

where

$$V_{m'm}^+ = e^{i\alpha_{m'm}\psi} e^{\beta_{-m',m}\phi}$$

and

$$V_{m'm}^- = e^{i\alpha_{m'm}\psi} e^{\beta_{m',-m}\phi}$$

In principle, nothing prevents us from factorizing the correlators (4.3) into two pieces:

$$\langle \prod_i V_{m'_i m_i}^{(\pm)}(z_i, \bar{z}_i) \rangle = \langle \prod_i e^{i\alpha_{m'_i m_i}\psi(z_i, \bar{z}_i)} \rangle \langle \prod_i e^{\beta_{m'_i m_i}^{\pm}\phi(z_i, \bar{z}_i)} \rangle \quad (4.4)$$

with

$$\beta_{m',m}^{\pm} = \begin{cases} \beta_{-m',m} \\ \beta_{m',-m} \end{cases}$$

Both correlators in (4.4) are standard correlators in CFT, and we can try to compute them using the standard techniques of the Coulomb gas formulation, *i.e.* using screening charges. The screening charges at our disposal are

$$e^{i\alpha_+\psi} \quad e^{i\alpha_-\psi} \quad e^{i\beta_+\phi} \quad e^{i\beta_-\phi}$$

As is standard in the Coulomb gas representation of correlators we will impose

$$\sum_i \alpha_{m'_i m_i} + n_+ \alpha_+ + n_- \alpha_- = \alpha_0 \quad (4.5)$$

$$\sum_i \beta_{m'_i m_i}^{\pm} + \tilde{n}_+ \beta_+ + \tilde{n}_- \beta_- = Q \quad (4.6)$$

where $n_{\pm}$ is the number of screening charges we must use in order to satisfy the charge neutrality condition (4.5). It is easy to check that the values of $\tilde{n}_{\pm}$ are determined by those of $n_{\pm}$:

$$\tilde{n}_+ = n_+ \quad \tilde{n}_- = -n_- - 1$$

Here we encounter the first major problem in the computation of the correlators (4.4), namely the need for a negative number of screening charges in the $c_L > 25$ sector. This

problem cannot be resolved by introducing “anti-screening” charges like $e^{-\beta-\phi}$ because, as can be easily checked with the help of equation (4.6), their conformal weights are not equal to one.

Recently, Dotsenko [23] introduced a trick to give meaning to the product of a negative number of screening charges. The identity

$$\prod_{i=a}^b f(i) = \frac{\prod_{i=1}^b f(i)}{\prod_{i=1}^{a-1} f(i)}$$

is valid when both a and b are positive integers. One may consider extending it to the case where b is negative, and thus

$$\prod_{i=1}^{-n-1} f(i) = \prod_{i=0}^n \frac{1}{f(-i)} \quad (4.7)$$

This very general trick can be applied successfully to the case at hand, and thus we can give meaning to any three-point correlation function. Before presenting the explicit computation, however, it might be worth pausing to discuss the difference between the operators V^+ and V_- appearing in (4.4).

Usually, in a (p', p) minimal model, we identify the primary fields labelled by (m', m) and $(p' - m', p - m)$ because their conformal weights are the same. In the Coulomb gas representation, these two primaries have charges $\alpha_{m'm}$ and $\alpha_{p'-m', p-m} = \alpha_0 - \alpha_{m'm} = \bar{\alpha}_{m'm}$: duality shows up in the Coulomb gas formalism through the symmetry $\alpha \leftrightarrow \alpha_0 - \alpha$.

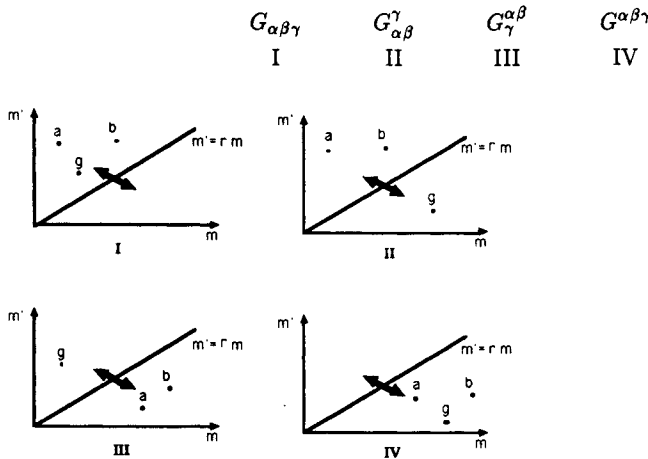
Taking into account this duplicity, given a value $\Delta_{m'm}$ for a conformal weight of the $c_M < 1$ model, we have at our disposal four different vertex operators of the full theory with total conformal weight equal to one:

$$\begin{aligned} V_{m'm}^+ &= e^{i\alpha_{m'm}\psi} e^{\beta_{m'm}^+\phi} \\ V_{m'm}^- &= e^{i\alpha_{m'm}\psi} e^{\beta_{m'm}^-\phi} \\ V_{m'm}^+ &= e^{i\bar{\alpha}_{m'm}\psi} e^{\beta_{m'm}^+\phi} \\ V_{m'm}^- &= e^{i\bar{\alpha}_{m'm}\psi} e^{\beta_{m'm}^-\phi} \end{aligned} \quad (4.8)$$

Do these four vertex operators represent the same physical state? To answer this question, we return to the BRST-based definition of physical state and observe that it was asymmetrical: we studied the cohomology on $\mathcal{F}^{(\text{ghost})} \otimes L_{m'm}(c_M) \otimes \mathcal{F}_{\beta, c_L}$,

with $\mathcal{F}_{\beta, c_L}$ the Fock module in the Coulomb gas representation of the theory $c_L > 25$. The asymmetry arises because for the $c_M < 1$ theory we already fix the Verma module $L_{m'm}(c_M)$. The two possibilities we see with this restriction are $V_{m'm}^+$ or $V_{m'm}^-$, depending on whether $\beta_{m'm}$ is larger or smaller than $Q/2$; the ghost number assignment is accordingly different. There is no reason to expect the other two vertex operators to yield any new states. From the cohomological point of view, thus, we find only two different states associated with the four vertex operators in (4.8), say $|\alpha_{m'm}; +\rangle$ and $|\alpha_{m'm}; -\rangle$. The label $\pm$ denotes the ghost number sign. To these states we may associate, furthermore, a “chirality” depending on whether the state is represented by a V^+ or a V^- operator. For instance, the state $|\alpha_{m'm}; -\rangle$ has $\beta < Q/2$ and its helicity is $(+)$ if $m' > m\rho$ and $(-)$ if $m' < m\rho$, with $\rho = \alpha_+/\alpha_-$. This description characterizes completely all the states. Note that we may change the chirality of a state by representing it with $\bar{\alpha}$ instead of α . Now we may impose the restriction $\beta \leq Q/2$, which means that the gravitational dressing is represented by a microscopic operator of the Liouville theory. The logic for this is, of course, to represent the external states entering into the scattering process by local operators. We will briefly comment on this condition in the last section.

Let us now review the results for the structure constants. To simplify the notation we shall use the following conventions. We label a generic three-point function with three indices, which are upper or lower depending on the chirality. The four different structure constants depend on the chiralities of the three intervening fields *i.e.* on what half of the chiral grid they lie (see figure):



The structure constants factorize in a piece corresponding to the theory $c_M < 1$ and another piece corresponding to the Liouville sector. Already at this level, an interesting

difference between the $c < 1$ and $c = 1$ theories can be observed. For the $c = 1$ theory processes with all external-state chiralities equal are kinematically forbidden. The rules of our computation are the following:

R1 . As far as the theory $c_1 < 1$ is concerned, the four cases I, II, III and IV above are identical, *i.e.* the c_1 part of the structure constants is the same. This is just like standard minimal models.

R2 . For the $c_1 < 1$ piece we use the *normalized* symmetric structure constants computed by Dotsenko and Fateev, *i.e.* we normalize $\langle \Phi_\alpha \Phi_{\alpha'} \rangle = \delta_{\alpha, \alpha'}$.

R3 . We perform all our computations for generic values of $\rho = \alpha_+/\alpha_-$ and only at the end do we take the rational limit $\rho \rightarrow p/p'$.

The rules R1 and R2 can be considered as the Coulomb gas expression of the BRST results quoted above. In fact, and on this point we disagree with Dotsenko's computation, we use for the matter part the exact structure constants of the minimal model for generic values of ρ , in the same way as for the BRST analysis the Virasoro module is defined on $L_{m'm} \otimes \mathcal{F}_{\beta, Q} \otimes \mathcal{F}^{(\text{ghost})}$ and not on $\mathcal{F}_{\alpha, \alpha_*} \otimes \mathcal{F}_{\beta, Q} \otimes \mathcal{F}^{(\text{ghost})}$.

We use the following vectors to characterize the screening content of the correlator $G_{\beta\gamma}^\alpha$:

$$(n_-, n_+) = \frac{\beta + \gamma - \alpha - 1}{2} \quad \text{for the } c_M \text{ model}$$

$$(\tilde{n}_-, \tilde{n}_+) = \left(\frac{-b' - d' + a' - 1}{2}, \frac{b + d - a - 1}{2} \right) \quad \text{for Liouville}$$

The generalization of this vector assignment to the other chirality assignments is obvious. The results are:

$$\begin{aligned} G_{\alpha\beta\gamma} &= K_{\alpha\beta\gamma} (a_\alpha a_\beta a_\gamma)^{1/2} \Delta(2 - \rho') \Delta(1 - \rho) \Delta_\alpha \Delta_\beta \Delta_\gamma \\ G_{\beta\gamma}^\alpha &= K_{\beta\gamma}^\alpha (a_\alpha)^{-1/2} (a_\beta a_\gamma)^{1/2} \Delta_{-\alpha} \Delta_\beta \Delta_\gamma \\ G^{\alpha\beta\gamma} &= K^{\alpha\beta\gamma} (a_\alpha a_\beta a_\gamma)^{1/2} \Delta(2 - \rho') \Delta(1 - \rho) \Delta^\alpha \Delta^\beta \Delta^\gamma \\ G_\alpha^{\beta\gamma} &= K^{-\alpha\beta\gamma} (a_\alpha)^{-1/2} (a_\beta a_\gamma)^{1/2} \Delta^{-\alpha} \Delta^\beta \Delta^\gamma \end{aligned} \quad (4.9)$$

where we have used the notations

$$\begin{aligned} \Delta(x) &= \frac{\Gamma(x)}{\Gamma(1-x)} \\ \rho &= \alpha_+/\alpha_- \\ \rho' &= \rho^{-1} \\ \Delta^\alpha &= \Delta(1 - a' + \rho a) \end{aligned}$$

$$K^{\alpha\beta\gamma} = \pi^{a'+b'+d'-2} \rho^{a'+b'+d'-a-b-d-2} \Delta(1-\rho)^{a+b+d} \quad (4.10)$$

$$a_\alpha^{-1} = C_{\alpha\alpha}^1 = \langle \Psi_\alpha \Psi_\alpha \Psi_1 \rangle \quad \text{for the } c_M \text{ model}$$

In the rational limit $\rho = p/p'$, we obtain the fusion rules for the minimal model but with truncation only from below. In the next section we shall see how to get the truncation from above by taking into account the ghost contribution to the correlators.

4.2 STRING AMPLITUDES AND GHOST CONTRIBUTIONS. TRUNCATION OF THE FUSION RULES

Amplitudes in string theory define measures on the moduli space of Riemann surfaces with punctures. The operator representation of these measures for the bosonic and fermionic string was presented in [27,28]. In this section we generalize some aspects of these measures to the case of non-critical strings.

We recall the main properties of the operator formalism in the case of the bosonic string. The operator formalism is based on interpreting a conformal field theory as a way to associate quantum states to Riemann surfaces. The Schrödinger representation of these states can be obtained using the functional integral representation and imposing boundary conditions at the punctures (see [27,28] for details and references). For the standard 26-dimensional bosonic string the two conformal field theories involved are the ghost system and 26 free bosonic fields. For non-critical strings the relevant conformal field theories contain the Liouville field together with the matter sector. We assume that at least formally it is possible to use the Liouville theory to associate states to Riemann surfaces using the functional integrals. The explicit form of these states will not be relevant for our general discussion.

Denote by $|\Sigma_{g,N}\rangle^{m,gh}$ the state associated to a Riemann surface of genus g and n punctures, where the superscript m denotes collectively the matter and Liouville fields and gh is the ghost contribution. The scattering measure for N external states $|\chi_1\rangle, \dots |\chi_N\rangle$ is given by

$$\langle \chi_1 | \langle \chi_2 | \dots \langle \chi_N | \prod_{i=1}^{3g-3+N} \bar{b}(v_i) b(v_i) | \Sigma_{g,n} \rangle^{(m)} | \Sigma_{g,n} \rangle^{(gh)} \quad (4.11)$$

where the $3g - 3 + N$ vector fields v_i represent a basis of the tangent space to the moduli space at the point given by the surface $\Sigma_{g,N}$.

The states $|\Sigma_{g,N}\rangle$ depend on what local coordinates we have chosen close to the punctures. A necessary condition for the amplitude (4.11) to define a good moduli measure is invariance under reparametrizations of the punctures. The transformation law of $|\Sigma_{g,N}\rangle$ under a change of coordinates $z \mapsto z + \epsilon v(z)$ is generated by the energy-momentum tensor $\oint T(z)v(z)dz$, with the integral taken around the puncture. Using this transformation law for the states $|\Sigma_{g,N}\rangle$, and taking into account that local changes of coordinates are generated by vector fields which are holomorphic in a neighborhood of the puncture and vanishing at its location, we obtain the following condition for the external states

$$L_n^{\text{tot}}|\chi_i\rangle = 0 \quad n \geq 0 \quad (4.12)$$

which are necessary for the reparametrization invariance of (4.11).

Defining the ghost number for the ghost fields $b(z)$ (resp. $c(z)$) as -1 (resp. $+1$), the ghost number of a generic state $|\Sigma_{g,N}\rangle$ is $3g - 3 + 3N/2$. In particular the state $|0\rangle$ associated to the Riemann sphere with one puncture has ghost number $-3/2$. This state is defined by conditions

$$b_n|0\rangle = 0 \quad n \geq -1 \quad c_n|0\rangle = 0 \quad n \geq 2 \quad (4.13)$$

These ghost number assignments imply for the external states appearing in (4.11)

$$\mathcal{G}|\chi_1\rangle \otimes |\chi_2\rangle \otimes \dots \otimes |\chi_N\rangle = -\frac{N}{2}|\chi_1\rangle \otimes |\chi_2\rangle \otimes \dots \otimes |\chi_N\rangle \quad (4.14)$$

From the point of view of the BRST cohomology, physical states are required to be in non-trivial cohomology classes. In the case of the critical bosonic string in flat Minkowski space (with signature $(25,1)$) and non-zero momentum for the center of mass, the only non-vanishing cohomology groups are H_m with $m = \pm 1/2$ [29]. The only way to saturate the constraint (4.14) is therefore with the ghost number of each of the external states in (4.11) taken as $-1/2$. To write explicitly the form of the states χ_i , we introduce two states associated to the ghost Hilbert space on the sphere : $|\pm\rangle$, with ghost numbers $\pm 1/2$ and satisfying the conditions

$$b_n|-\rangle = 0 \quad n \geq 0 \quad c_n|+\rangle = 0 \quad n \geq 0 \quad (4.15)$$

These states satisfy $L_0|\pm\rangle = -1|\pm\rangle$; hence the state χ_i becomes $|\chi_i\rangle = |\chi_i\rangle^{(m)} \otimes |-\rangle$. For these states the BRST condition $Q|\chi\rangle = 0$ implies the condition (4.12).

Now we are ready to extend the previous result to the case of non-critical strings. The main change comes from the BRST cohomology structure discussed in section 2.2. Unlike the critical string case, there are BRST cohomology classes with ghost numbers not restricted to being $\pm 1/2$ (cf. (3.31)). The non-vanishing classes are labelled by

$$H_{d_\beta \pm 1/2} \quad \beta \leq Q/2$$

$$H_{-d_\beta \pm 1/2} \quad \beta > Q/2 \quad (4.16)$$

Recall that d_β is the distance from the tip in the tower of embeddings (3.30) of the state chosen. The condition (4.14) is however independent of whether we work with critical or non-critical strings. Its origin is purely geometrical. Therefore if we restrict ourselves to microscopic states i.e. $\beta \leq Q/2$, the only possible non-vanishing amplitudes are those restricted to the standard dressed primary fields of the minimal CFT $\Psi_{m',m}$, $1 \leq m \leq p-1$, $1 \leq m' \leq p'-1$ or its boundaries of the form $(o, m), (m', o)$. We will call this domain in the Kac table the minimal conformal grid. The theorem of Lian and Zuckerman was strictly proved for fields in the minimal grid; however there does not seem to be any obstruction to extending it to boundary states, and we will assume this to be the case. All we need is the ghost number of the operator $\Psi_{m',0}$. According to the theorem of Lian and Zuckerman, these fields all have ghost number equal to $-1/2$ while the other fields associated to the null vectors of the CFT and with $\beta \leq Q/2$ always have ghost number $\geq 1/2$. Therefore the correlators involving two fields in the minimal grid and one of the null descendants do not satisfy (4.14), and they vanish. A similar argument applies to the three-point function involving two null descendants and one ordinary field. More precisely we can say that any correlator involving any (different from zero) number of null descendants satisfying the condition $\beta \leq Q/2$ vanishes identically on any genus Riemann surface as a consequence of ghost number counting.

We now reconsider the study of the fusion rules started in the previous section. There, it was shown using Dotsenko's prescription [23] that one is led generically to the irrational fusion rules. If we apply these fusion rules we are sometimes driven outside the minimal grid. However, the ghost-induced decouplings truncate the fusion rules to:

$$1 + |\beta - \delta| \leq \alpha \leq \min\{\beta + \delta - 1, \kappa\}, \quad \alpha + \beta + \delta = \text{odd} \quad (4.17)$$

where $\kappa \equiv (p', p)$. These are not exactly the standard fusion rules, but only the irrational ones restricted to the minimal conformal grid. It is remarkable that (4.17)

do not give an associative OPE. So the ghosts appear to spoil the associativity of the irrational OPE, as inherited from the matter sector. The only associative truncation of the irrational fusion rules yields the standard ones.

As a matter of fact, using the enhanced symmetry of the rational theories, one can choose screening conventions (i.e. renormalization prescriptions) such that the standard ones are obtained. The point is that the Liouville "energies" (β), only depend on the combination $|np' - n'p|$; thus they are invariant under the symmetry $\alpha = (n', n) \rightarrow \alpha^* = (p' - n', p - n)$ (in this case $\alpha^* \equiv \bar{\alpha}$ because $p\alpha_+ + p'\alpha_- = 0$). In consequence we can avoid the chiral distinction for rational theories and set up the prescription:

$$L_{\alpha\beta\delta} = L_{\alpha}^{\beta^*\delta^*} \quad (4.18)$$

ensuring that the Liouville part is insensitive to the $\star$ transformation. If we select now the screening convention for matter as the one involving a minimal number of screenings, then we truncate to BPZ's:

$$1 + |\beta - \delta| \leq \alpha \leq \min\{\beta + \delta - 1, 2\kappa - 1 - \beta - \delta\} \quad \alpha + \beta + \delta = \text{odd} \quad (4.19)$$

In this way the standard fusion rules are recovered.

4.3 FINAL COMMENTS

We would like to end this lectures with some general comments.

The space-time interpretation as non-critical strings of conformal field theories coupled to gravity is that of a two-dimensional target space-time with non-vanishing backgrounds for the dilaton and tachyon field. If, on the other hand, we consider the string obtained by choosing as matter a conformal field theory with central extension 26 resulting of combining two Feigin-Fuchs fields, the space-time picture we obtain is again a two-dimensional target space-time but with vanishing tachyon background. *A priori*, both possibilities are on equal footing and certainly both are solutions to the constraint that the β -function equations vanish. Why nature would choose the Liouville case instead of the simpler one corresponding to two Feigin-Fuchs fields is a dynamical non-perturbative question.

As we have discussed in the first lecture, some Liouville correlators in the classical limit are good candidates for solutions to the uniformization equation. In this sense, it would be very interesting to check this conjecture for the correlators obtained by analytic continuation techniques using as puncture operator $\phi e^{\gamma\phi}$.

Another open problem is to find a sewing procedure compatible with the computation of correlators performed with analytic continuation. It should be expected that only macroscopic states contribute to it.

In the free field approach to Liouville theory used in the last lecture, the difference between macroscopic and microscopic operators lost part of its original physical meaning. In fact, nothing prevents us from computing, using the analytic continuation technology, correlators in which macroscopic operators appear. In principle, in the free field approach, the only difference between both kind of states is that they belong to different BRST cohomology classes. It would be very interesting to prove, using this BRST information, that only when we respect the constraint $\beta \leq Q/2$, the correlations define good measures on moduli space invariant under general reparametrizations.

Acknowledgements. We thank J.L.F. Barbón, M. Ruiz-Altaba, G. Sierra and A.B. Zamolodchikov for interesting conversations.

REFERENCES

- [1] N. Seiberg: *Notes on Quantum Liouville Theory and Quantum Gravity*. RU-90-29. Talk presented at the 1990 Yukawa International Seminar.
- [2] V.G. Knizhnik, A.M. Polyakov and A.B. Zamolodchikov: *Mod. Phys. Lett.* **A3** (1988) 819.
- [3] M. Bershadsky and I. Klebanov: PUPT-1241 (1991).
- [4] A.A. Belavin, A.M. Polyakov and A.B. Zamolodchikov: *Nucl. Phys.* **B241** (1984) 333.
- [5] D. Friedan, Z. Qiu and S. Shenker: *Phys. Rev. Lett.* **51** (1984) 1575.
- [6] A.M. Polyakov : *Phys. Lett.* **103 B** (1981) 207.
- [7] H. Farkas and I. Kra : *Riemann Surfaces*. Springer Verlag, 1984.
- [8] J.A. Hempel: *Bulletin of London Math. Society* **20** (1988) 97.
- [9] V.S. Dotsenko and V.A. Fateev: *Nucl. Phys.* **B240** (1984) 312; *Nucl. Phys.* **B251** (1985) 691; *Phys. Lett* **154B** (1985) 291.
- [10] D. Zograf and L. Takhtadjan: *Russ. Math. Surveys* **42** (1987) 169; *Math. USSR Izv.* **35**(1) (1990) 83. L. Takhtadjan: *Lectures on Theta Functions* (American Mathematical Society 1989).
- [11] T. Curtright and C.B. Thorn, *Phys. Rev. Lett.* **48** (1982) 1309. See also E. Braaten, T. Curtright and C.B. Thorn, *Phys. Lett.* **118 B** (1982) 115, *Ann. Phys.* **147** (1983) 365; E. Braaten, T. Curtright, G. Ghandour and C.B. Thorn, *Phys. Rev. Lett.* **51** (1983) 19; *Ann. Phys.* **153** (1984) 147.
- [12] J.L. Gervais and A. Neveu: *Nucl. Phys.* **B199** (1982) 59, *Nucl. Phys.* **B209** (1982) 125, *Nucl. Phys.* **B224** (1983) 329, *Nucl. Phys.* **B238** (1984) 125, 396; *Phys. Lett.* **151B** (1985) 271.
- [13] J.L. Gervais: *Comm. Math. Phys.* **130** (1990) 257; *Comm. Math. Phys.* **138** (1990) 301. F. Smirnov and L. Takhtadjan : *Towards an Quantum Liouville Theory for $c > 1$* . Boulder, Colorado Preprint, 1990. C. Gomez and G. Sierra: *Phys. Lett.* **255B** (1991) 51.
- [14] B.L. Feigin and D.B. Fuks, Moscow Preprint 1983.
- [15] For details and references see for example: A. Rocha in the Proceedings of the 10th John Hopkins Workshop on Current Problems in Particle Theory. V. Rittenberg ed. World Scientific, Singapore 1986.

- [16] V.S. Dotsenko and V.A. Fateev: Nucl. Phys.**B240** (1984) 312; Nucl. Phys.**B251** (1985) 691; Phys. Lett **154B** (1985) 291.
- [17] J. Distler and H. Kawai: Nucl. Phys.**B321** (1989) 509.
- [18] F. David: Mod. Phys. Lett. **A3** (1988) 1651.
- [19] B. Lian and G. Zuckerman: Phys. Lett. **B 254** (1990) 417.
- [20] M. Goulian and M. Li: Phys. Rev. Lett. **66-16** (1991) 2051.
- [21] Y. Kitazawa: HUTP-91/A013.
- [22] P. Di Francesco and D. Kutasov: Phys. Lett. **261B** (1991) 385.
- [23] V.S. Dotsenko: PAR-LPTHE 91-18.
- [24] L. Alvarez-Gaumé, J.L.F. Barbón and C. Gómez: *Fusion Rules in Two-Dimensional Gravity*. CERN-TH-6142/91.
- [25] A.M. Polyakov: Mod. Phys. Lett. **A 6** (1990) 635.
- [26] E. Martinec, G. Moore and N. Seiberg: RU-14-91, YCTP-P10-91, EFI-91-14
- [27] L. Alvarez-Gaumé, C. Gómez, G. Moore and C. Vafa: Nucl. Phys.**B303** (1988) 455.
- [28] L. Alvarez-Gaumé, C. Gómez, P. Nelson, G. Sierra and C. Vafa: Nucl. Phys.**B311** (1988/89) 333.
- [29] I. Frenkel, H. Garland and G. Zuckerman; Proc. Natl. Acad. Sci. USA **83** (1986) 8442.

BLACK HOLES AND STRINGS IN TWO DIMENSIONS*

HERMAN VERLINDE†

*Joseph Henry Laboratories**Princeton University**Princeton, NJ 08544*

ABSTRACT

In the first part of these notes some properties of string theory in the two dimensional black hole target space are discussed. In the second part we consider the quantization of the effective target space field theory, which provides an interesting toy model for studying black hole evaporation due to Hawking radiation.

*Lectures given at the ICTP Spring School on Strings and Quantum Gravity, Trieste, April 1991, and at the 6th M.Grossmann Meeting, Kyoto, June 1991.

†Supported by NSF Grant PHY90-21984

1. INTRODUCTION

Black hole physics is notorious for its many paradoxes, some of which even challenge the fundamental principles of classical and quantum physics. Important questions such as those concerning the nature of the singularity or the ultimate fate of an evaporating black hole are still unanswered, the main obstacle being the lack of a consistent, manageable theory of quantum gravity. It is hoped that string theory will ultimately provide a natural resolution to these problems, but it is clear that to properly address them in a general setting its formalism needs to be better understood. Two surprising recent developments, however, have opened up the possibility to gain important insight into these questions in the context of string theory in two target space dimensions. While on the one hand our understanding of two-dimensional string theory has improved dramatically in the past year due to the success of the matrix model techniques [1], it was recently observed by Witten [3] that it in fact admits a black hole geometry as its target space.

Two dimensional string theory provides a very interesting theoretical laboratory to learn more about the fundamental difference between string theory and conventional quantum field theory. Namely, when a string propagates in a $1+1$ dimensional space-time, there are no transversal directions. So the string can not vibrate and is classically completely described by its center of mass coordinates. Many of the usual distinctions between strings and point-particles therefore disappear, and $1+1$ dimensional string theory indeed contains essentially only one field theoretic degree of freedom. The general sigma model describing the world sheet action for strings in a $1+1$ dimensional target space, parametrized by coordinates $x = (x^1, x^2)$, has the form

$$S = \frac{1}{2\pi} \int_{\Sigma} \left(G_{ij}(x) \bar{\partial} x^i \partial x^j + \Phi(x) R^{(2)} + T(x) \right), \quad (1.1)$$

where $G_{ij}(x)$, and $\Phi(x)$ are the space-time metric and dilaton field and $T(x)$ is the tachyon. The dynamics of the background fields is governed by the condition that the β -functions of the worldsheet field theory vanish. These equations are encoded in the target space action [2]

$$S_{eff} = \int d^2x \sqrt{G} e^{\Phi} \left(R + (\nabla \Phi)^2 + (\nabla T)^2 - 2T^2 - 8 \right). \quad (1.2)$$

The tachyon $T(x)$ in fact describes a massless field. The metric and dilaton field (and all the other massive modes) only have global modes, and serve as a background for the propagating tachyon.

It was shown in [3, 4] that if one puts $T = 0$ the classical equations of motion for the metric and dilaton admit the following solution (here $x^i = (u, v)$)

$$ds^2 = \frac{du dv}{1 - uv}, \quad (1.3)$$

$$\Phi = \log(1 - uv). \quad (1.4)$$

The target space geometry described by this classical metric is that of a two-dimensional black hole (see fig. 1). The horizon is located at $uv = 0$, and the future and past singularities at $uv = 1$. The regions *I* and *II* describe the space-time outside of the horizon, whereas regions *III* and *IV* correspond to the future and past interior of the horizon. The following analytic continuation can be performed in region *I* where we can express u and v in terms of a radial coordinate r and a time coordinate t via

$$u = \sinh r e^t, \quad v = -\sinh r e^{-t} \quad (1.5)$$

The metric and dilaton become

$$ds^2 = dr^2 - \tanh^2 r dt^2, \quad (1.6)$$

$$\Phi = \log \cosh^2 r. \quad (1.7)$$

These (r, t) coordinates are similar to Schwarzschild coordinates, while u and v are the analogs of the Kruskal coordinates.

The first part of these notes summarizes work [6] done in collaboration with R. Dijkgraaf and E. Verlinde on the propagation of strings in this black hole background. Starting point for this study will be the observation of Witten [3] that the sigma model described by (1.3) can in fact be represented as an exactly soluble conformal field theory, obtained

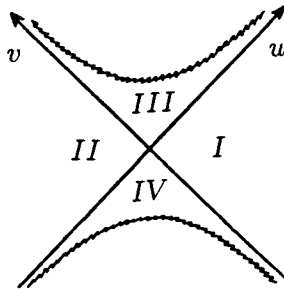


Fig. 1: The two-dimensional black hole in (u, v) -coordinates.

from the $SL(2, R)$ WZW-model by dividing out a (non-compact) $U(1)$ sub-group via the GKO coset construction. We will make use of this CFT technology to compute the scattering amplitude of a string off the black hole, and to discuss some of the remarkable duality properties of the Euclidean black hole solution.

As mentioned above, one of the most intriguing aspects of black hole physics is that a black hole is not stable but emits thermal radiation [15]. Clearly it would be interesting if one could explicitly study the resulting black hole evaporation process in the two-dimensional string model. Unfortunately, however, the present formalism of string theory is not yet powerful enough, since for this we would in particular need a second quantized formalism which in addition is capable of describing the gravitational back reaction due to the outflux of matter. On the other hand, since black hole evaporation is already poorly understood in ordinary field theory, as a first step we could in fact try to describe it in terms of the effective low energy field theory (1.2). In the last part of these notes we will make some remarks concerning the quantization of the low energy field theory, which may perhaps become helpful in future investigations in this direction.

2. STRING THEORY IN THE BLACK HOLE BACKGROUND.

In this section we describe the space of primary operators of the $SL(2, R)/U(1)$ coset conformal field theory and discuss their interpretation in terms of string scattering in the black hole background.

2.1. THE $SL(2, R)/U(1)$ COSET MODEL

The black hole conformal field theory is obtained from the WZW-model based on the non-compact group $G = SL(2, R)$, by gauging the symmetry

$$g \rightarrow hgh, \quad (2.1)$$

with h in an appropriately chosen abelian subgroup H of $G = SL(2, R)$. If one chooses H to be compact, the coset manifold G/H describes a Euclidean version of the two-dimensional black hole, while the construction with H non-compact leads to a black hole target space of Lorentzian signature. The two constructions are related to each other via analytic continuation.

The coordinates u and v in the Minkowskian black hole metric (1.3) correspond to the parametrization of the $SL(2, R)/U(1)$ coset as [3]

$$g(u, v) = \begin{pmatrix} a & u \\ -v & b \end{pmatrix}, \quad ab + uv = 1, \quad (2.2)$$

where a and b are redundant variables under the gauge symmetry $g \rightarrow hgh$.^{*} Another convenient set of coordinates r, t_L, t_R are the familiar Euler angles of $SL(2, R)$

$$g = e^{\frac{i}{2}t_L\sigma_3} e^{\frac{1}{2}r\sigma_1} e^{\frac{i}{2}t_R\sigma_3} \quad (2.3)$$

with σ_i the Pauli matrices and the coordinates range over $0 \leq r < \infty$, $-\infty < t_{L,R} < \infty$. The generator of the abelian subgroup H is taken to be σ_3 , so that the local gauge transformations correspond to shifting $t_{L,R} \rightarrow t_{L,R} + \alpha$.

^{*}Strictly speaking, the $SL(2, R)/U(1)$ coset is a double cover of the (u, v) plane, where the two sheets are distinguished by the sign of a or b . The two sheets intersect at the singularity $uv = 1$.

In the parametrization (2.3) the gauged WZW action takes the form

$$S = S_{wzw}[r, t_L, t_R] + \frac{k}{2\pi} \int d^2 z \left[A(\bar{\partial} t_R + \cosh r \bar{\partial} t_L) + \bar{A}(\partial t_L + \cosh r \partial t_R) + \bar{A}A(\cosh r + 1) \right] \quad (2.4)$$

with

$$S_{wzw}[r, t_L, t_R] = \frac{k}{4\pi} \int d^2 z (\bar{\partial} r \partial r + \bar{\partial} t_L \partial t_L + \bar{\partial} t_R \partial t_R + 2 \cosh r \bar{\partial} t_L \partial t_R) \quad (2.5)$$

To analyze the spectrum of this theory, we will impose the Lorentz gauge condition

$$\partial_\alpha A^\alpha = 0, \quad (2.6)$$

Assuming that the world sheet has trivial topology, we may parametrize the gauge slice via

$$A^\alpha = \epsilon^{\alpha\beta} \partial_\beta \phi \quad (2.7)$$

The gauged fixed action then attains the following form

$$S_{gf} = S_{wzw}[r, t_L + \phi, t_R - \phi] + S[\phi] + S[b, c] \quad (2.8)$$

where

$$S[\phi] = -\frac{k}{4\pi} \int d^2 z \partial \phi \bar{\partial} \phi \quad (2.9)$$

describes a (wrong sign) free scalar field and

$$S[b, c] = \int d^2 z (b \bar{\partial} c + \bar{b} \partial c), \quad (2.10)$$

describes a spin (1, 0) ghost system. The quantization of the gauge fixed theory (2.8) is straightforward, since everything is expressed either in free fields, or in terms of an (ungauged) $SL(2, R)$ WZW-model. The operators that are part of the coset theory are defined in terms of the BRST-symmetry of the combined system (2.8): one projects the Hilbert space of the full theory on the invariant subsector and identifies states whose difference is BRST-exact. We will not explicitly describe this procedure here, but directly turn to the description of the resulting space of primary operators of the coset theory. For more details we refer to [6].

2.2. PRIMARY VERTEX OPERATORS.

The primary fields of the coset chiral algebra represent the vertex operators for the tachyon field. The other physical vertex operators given by Virasoro primary fields which are descendents of the coset chiral algebra will not be considered here (but see *e.g.* [9] and [10]). The most general Ansatz for the coset primary fields is that they are given by arbitrary functions $T(g)$ on $SL(2, R)$, invariant under the gauge symmetry

$$T(hgh) = T(g) \quad (2.11)$$

It is a well-known fact that a complete basis of functions on a group consists of the matrix elements of the different representations. In our situation we will be interested in the unitary representations of the universal cover $\widetilde{SL}(2, R)$. The vectors $|l, \omega\rangle$ in these infinite dimensional representations are labeled by the eigenvalue ω of the generator of the abelian subgroup H . According to the values of the $SL(2, R)$ isospin l and ω the unitary representation can be divided into three different series:

- principal continuous series: $l = -\frac{1}{2} + i\lambda$; λ, ω real;
- principal discrete series: $l < -\frac{1}{2}$; $|\omega| + l$ non-negative integer;
- complementary series: $l \in [-1, -\frac{1}{2}]$, ω real.

We can now choose as a basis of functions $T(g)$ the matrix elements

$$T(g) = \langle l, \omega_L | g | l, \omega_R \rangle. \quad (2.12)$$

The quantum numbers ω_L and ω_R are the eigenvalues of the left- and right-moving charges J_0^3 and $\bar{J}_0^3$ respectively. Gauge invariance implies that

$$\omega_L = -\omega_R = \omega \quad (2.13)$$

It turns out that with this condition the matrix elements (2.12) are only non-zero for the continuous series with $l = -\frac{1}{2} + i\lambda$. Hence the primary vertex operators take the form

$$T_\omega^l(r, t) = \langle l, \omega | g(r, t) | l, -\omega \rangle, \quad (2.14)$$

where $l = -\frac{1}{2} + i\lambda$ and $t = \frac{1}{2}(t_L - t_R)$ is the time-variable.

The stress tensor of the G/H coset model is given by the difference of the G and H Sugawara stress tensors

$$T(z) = \frac{1}{k-2} \eta_{ab} J^a J^b - \frac{1}{k} (J^3)^2 \quad (2.15)$$

where η_{ab} denotes the metric on the lie algebra $sl(2, R)$. If $k = 9/4$ the central charge is $c = 26$, and the coset theory can be used to describe a critical string theory in, as we will see, the two-dimensional black hole background (1.3). The modes of the stress tensor $T(z)$ generate the Virasoro algebra with central charge

$$c = \frac{3k}{k-2} - 1$$

Now, we observe that of the positive modes J_n^a with $n \geq 0$ of the $SL(2, R)$ Kac-Moody currents only the zero-modes J_0^a act non-trivially on the tachyon vertex operators, and can be represented as the differential operators. From this we derive that the Virasoro operator L_0 can be identified with a following wave operator on the (r, t) target space

$$L_0 = -\frac{\Delta_0}{k-2} + \frac{1}{k} \frac{\partial^2}{\partial t^2} \quad (2.16)$$

where Δ_0 is the Casimir operator on $SL(2, R)$ restricted to H-invariant functions:

$$\Delta_0 = \frac{\partial^2}{\partial r^2} + \coth r \frac{\partial}{\partial r} + \frac{1}{\sinh^2 \frac{1}{2}r} \frac{\partial^2}{\partial t^2} \quad (2.17)$$

As matrix elements, the vertex operators (2.14) are simultaneous eigenfunctions of $\partial/\partial t$ and the Casimir operator Δ_0 . Hence they are eigen modes of the above wave operator with eigenvalues

$$L_0 |T_\omega^l\rangle = \left(-\frac{l(l+1)}{k-2} - \frac{\omega^2}{k} \right) |T_\omega^l\rangle, \quad (2.18)$$

Notice that for all the unitary representations the Casimir $l(l+1)$ is real, but that only for the continuous series $-l(l+1) = \lambda^2 + \frac{1}{4} > 0$. The parameter λ has the interpretation of spatial momentum, whereas ω corresponds to the energy of the string state. In the critical case, $k = 9/4$, the mass shell condition reads

$$(L_0 - 1) |T_\omega^\lambda\rangle = (4\lambda^2 - \frac{4}{9}\omega^2) |T_\omega^\lambda\rangle = 0 \quad (2.19)$$

The two solutions $\lambda = \pm\omega/3$ describe the incoming and outgoing tachyon modes.

2.3. REPRESENTATIONS AND MATRIX ELEMENTS

To obtain a more explicit description of the tachyon vertex operators we will need to borrow some results from the representation theory of $SL(2, R)$. The representations relevant to us will be the infinite dimensional unitary ones. These can be constructed by considering the natural action of group $SL(2, R)$ on functions $f(x, y)$ on the plane

$$g : f(x, y) \rightarrow f(ax + by, cx + dy). \quad (2.20)$$

The so-called principal continuous series of representations is obtained if we demand that f scales as

$$f(ax, ay) = |a|^{2l} (\text{sgn } a)^{2\epsilon} f(x, y). \quad (2.21)$$

So, a particular representation is labeled by l , a complex number that represents the $SL(2, R)$ spin, and $\epsilon = 0, \frac{1}{2}$ indicating whether f has even or odd parity:

$$f(-x, -y) = (-1)^{2\epsilon} f(x, y). \quad (2.22)$$

Through this scaling f effectively depends only on one real variable, and we can write *e.g.* $f(x) = f(x, 1)$. We will now construct the matrix elements of these representations. To this end we have to choose an abelian subgroup H to label our states. In our case H is non-compact and the orbits are hyperbolas in the (x, y) plane. If one uses $SL(2, R)$ notation, H can be chosen of the form

$$h = \begin{pmatrix} e^{\frac{1}{2}t} & 0 \\ 0 & e^{-\frac{1}{2}t} \end{pmatrix}, \quad -\infty < t < \infty. \quad (2.23)$$

An orbit consists of four disconnected parts. A function $f(x, y)$ in a particular representation is completely determined by its values on the line $y = 1$. So, we can work with the functions $f(x) = f(x, 1)$, that transform as

$$g : f(x) \rightarrow |cx + d|^{2l} \text{sgn}^{2\epsilon}(cx + d) f\left(\frac{ax + b}{cx + d}\right). \quad (2.24)$$

The eigenstates of H , with eigenvalue ω are given by

$$f_{\omega}^{\pm}(x) = x^{i\omega + l} \theta(\pm x). \quad (2.25)$$

Similarly as in the compact case, matrix elements can be expressed as integrals over the real line

$$\pm \langle l, \omega_L | g | l, \omega_R \rangle_{\pm} = \int \frac{dx}{2\pi i x} f_{\omega_L}^{\pm}(x) (g \cdot f_{\omega_R}^{\pm}(x)) \quad (2.26)$$

Depending on the group element g and the $\pm$ signs the integral ranges over some specific interval. In this way we find the representation

$$T_{\omega}^{\lambda}(u, v) = \int_{\mathcal{C}} \frac{dx}{x} x^{-2i\omega} (\sqrt{1-uv} + u/x)^{-\nu_-} (\sqrt{1-uv} - vx)^{-\nu_+} \quad (2.27)$$

where

$$\nu_{\pm} = \frac{1}{2} - i(\lambda \pm \omega). \quad (2.28)$$

In fact, there are several possible choices for the contour of integration $\mathcal{C}$ since the integrand has a number of branch cuts, at $x = 0$ and $x = \infty$, and two other ones at

$$x_1 = \sqrt{1-uv}/v, \quad \text{and} \quad x_2 = -u/\sqrt{1-uv}. \quad (2.29)$$

This ambiguity in the choice for $\mathcal{C}$ is directly related to the fact that one actually has two eigenstates for fixed ω in the principal continuous series of $SL(2, R)$ representations, and correspondingly four different vertex operators with the quantum numbers of T_{ω}^{λ} .

2.4. SCATTERING OFF THE BLACK HOLE

Although a classical particle would simply fall into the black hole, a quantum mechanical wave can scatter back. The vertex operators (2.27) can be used to describe this quantum scattering of strings off the black hole. For this it is sufficient to consider the region I , in which $u > 0$, $v < 0$. In this case the two branch points x_1 and x_2 in (2.27) are located on the negative real axis. Together with the branch point at $x = 0$ they divide up the real axis in four segments $\mathcal{C}_i$ ($i = 1, \dots, 4$), each of which can be taken as our contour $\mathcal{C}$. In this way we find the following four types of vertex operators

$$\begin{aligned} R_{\omega}^{\lambda}(u, v) &: \text{ for } \mathcal{C}_1 = [0, \infty[, \\ U_{\omega}^{\lambda}(u, v) &: \text{ for } \mathcal{C}_2 = [x_2, 0], \\ L_{\omega}^{\lambda}(u, v) &: \text{ for } \mathcal{C}_3 = [x_1, x_2], \\ V_{\omega}^{\lambda}(u, v) &: \text{ for } \mathcal{C}_4 =]-\infty, x_1]. \end{aligned} \quad (2.30)$$

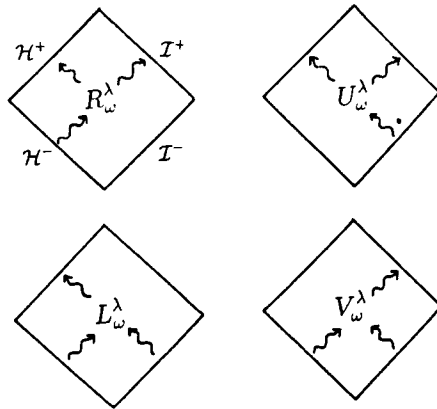


Fig. 2: The behaviour of the four modes R_ω^λ , L_ω^λ , U_ω^λ , and V_ω^λ (defined in the text) at the infinities of region I . Each mode vanishes at one particular null boundary.

All these contour integrals can be evaluated in terms of the gaussian hypergeometric function $F(\alpha, \beta; \gamma; z)$. The physical interpretation of the four different modes becomes clear when we consider their behaviour at spatial infinity ($r = \infty$) and at the horizon ($r = 0$). The asymptotic behaviour of the operators can be read off immediately from the integral representation. One finds that in the limit $r \rightarrow \infty$ we have in (r, t) coordinates[†]

$$\begin{aligned} L_\omega^\lambda &\sim e^{-\frac{1}{2}r} e^{-i\lambda r - 2i\omega t} \\ R_\omega^\lambda &\sim e^{-\frac{1}{2}r} e^{i\lambda r - 2i\omega t} \end{aligned} \quad (2.31)$$

Hence the asymptotic forms of the modes L_ω^λ and R_ω^λ represent respectively pure left-moving and right-moving tachyon waves at spatial infinity. A similar analysis shows that the modes U_ω^λ and V_ω^λ describe pure left or right-moving plane waves in the neighbourhood of the horizon $r = 0$

$$\begin{aligned} U_\omega^\lambda &\sim u^{-2i\omega} \\ V_\omega^\lambda &\sim (-v)^{2i\omega} \end{aligned} \quad (2.32)$$

Hence U and V become infinitely blue-shifted when they approach the past resp. future horizon. The vertex operator U_ω^λ describes a tachyon falling into the future horizon, that

[†]Here we assume that the mass-shell condition (2.19) is satisfied, and that the energy ω is positive.

is, the absorption of the string mode by the black hole. On the other hand the operator V_ω^λ describes the process where a particle is emitted by the 'white hole' and crosses the past horizon. Summarizing, we see that wave packets constructed out just one of the four possible modes will vanish at one particular null infinity, as illustrated in fig. 2.

Let us now calculate the amplitude for the back-scattering of a tachyon off the black hole geometry. To describe this physical process we need a mode that represents only absorption and no emission of particles by the black hole, and which therefore vanishes at the past horizon $\mathcal{H}^-$. From the given asymptotic behaviour we see that the correct mode is U_ω^λ , which has no component propagating out of the horizon. The reflection coefficient $S(\lambda, \omega)$ can now be determined by decomposing U_ω^λ in plane waves at $r \rightarrow \infty$, far away from the black hole

$$U_\omega^\lambda \sim e^{-\frac{1}{2}r - 2i\omega t} (e^{-i\lambda r} + S(\lambda, \omega) e^{i\lambda r}) \quad (2.33)$$

From the integral representation (2.27-2.30) of U_ω^λ we read off that

$$S(\lambda, \omega) = \frac{\cosh \pi(\lambda - \omega) B(\nu_+, \nu_-)}{\cosh \pi(\lambda + \omega) B(\bar{\nu}_+, \bar{\nu}_-)} \quad (2.34)$$

where $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ denotes the Euler β -function. Inserting the mass-shell condition $\omega = 3\lambda > 0$, we get

$$S(\lambda) = \frac{\Gamma(1 + 2i\lambda)\Gamma^2(\frac{1}{2} - 4i\lambda)}{\Gamma(1 - 2i\lambda)\Gamma^2(\frac{1}{2} - 2i\lambda)} \quad (2.35)$$

We note that the reflection coefficient is not a phase factor, but

$$|S(\lambda)| = \frac{\cosh 2\pi\lambda}{\cosh 4\pi\lambda} \quad (2.36)$$

This is in accordance with the expectation that only part of the tachyon wave gets reflected, the other part will enter the horizon and will be absorbed by the black hole.

3. DUALITY AND THE EUCLIDEAN BLACK HOLE CFT

The Euclidean black hole metric is obtained from the Minkowskian one (1.6) by replacing $t \rightarrow i\theta$.

$$ds^2 = dr^2 + \tanh^2 r d\theta^2, \quad (3.1)$$

This metric describes a regular manifold having the shape of an semi-infinite cigar when the θ -coordinate is chosen to be periodic modulo 2π (see fig. 3). In this section we will review some of its properties. In this case we will discover some more stringy phenomena, such winding states and target space duality.

Also the Euclidean black hole allows an exact CFT description in terms of a $SL(2, R)$ mod $U(1)$ coset model [3], where now the $U(1)$ is compact. The coordinates r and θ are related to the Euler angles of $SL(2, R)$

$$g = e^{\frac{i}{2}\theta_L\sigma_3} e^{\frac{1}{2}r\sigma_1} e^{\frac{i}{2}\theta_R\sigma_3} \quad (3.2)$$

The vertex operators in the Euclidean coset theory are still given by the matrix elements of g

$$\begin{aligned} T_{mn}^l(r, \theta_L, \theta_R) &= \langle \omega_L | g | \omega_R \rangle \\ &= \mathcal{P}_{\omega_L \omega_R}^l(\cosh r) e^{i\omega_L \theta_L + i\omega_R \theta_R} \end{aligned}$$

where the functions $P_{\omega, \omega'}^l(x)$ are the so-called Jacobi functions [11]. The allowed values of ω_L and ω_R are

$$\omega_L = \frac{1}{2}(m + nk), \quad \omega_R = -\frac{1}{2}(m - nk), \quad (3.3)$$

with n and m integers. The quantum number m is to be interpreted as the discrete momentum of the string in the θ direction, while n is the winding number. The vertex operators (3.3) seem to depend on three independent fields, however, because the WZW-theory is gauged, the θ fields effectively satisfy the constraints $\partial\theta_L + \cosh r \partial\theta_R = \bar{\partial}\theta_R + \cosh r \bar{\partial}\theta_L = 0$. Therefore it is more appropriate to think of θ_L and θ_R as the left- and right-moving parts of a *single* coordinate θ .

Our aim will be to extract from the above tachyon spectrum information about the geometry of the Euclidean black hole. To this end, let us again consider the L_0 -operators.

$$L_0 = -\frac{\Delta_0}{k-2} - \frac{1}{k} \frac{\partial^2}{\partial \theta_L^2},$$

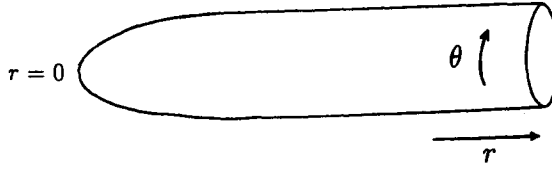


Fig. 3: The Euclidean black hole geometry in which the momentum modes of the tachyon propagate, as parametrized by the coordinates (r, θ) . This geometry is regular at the point $r = 0$.

$$\bar{L}_0 = -\frac{\Delta_0}{k-2} - \frac{1}{k} \frac{\partial^2}{\partial \theta_R^2}, \quad (3.4)$$

where the Casimir Δ_0 is now

$$\Delta_0 = \frac{\partial^2}{\partial r^2} + \coth r \frac{\partial}{\partial r} + \frac{1}{\sinh^2 r} \left(\frac{\partial^2}{\partial \theta_L^2} - 2 \cosh r \frac{\partial^2}{\partial \theta_L \partial \theta_R} + \frac{\partial^2}{\partial \theta_R^2} \right) \quad (3.5)$$

The matrix elements T_{mn}^l are simultaneous eigenfunctions of L_0 and $\bar{L}_0$, with eigenvalues

$$L_0 |T_{mn}^l\rangle = h_{mn}^l |T_{mn}^l\rangle, \quad \bar{L}_0 |T_{mn}^l\rangle = \bar{h}_{mn}^l |T_{mn}^l\rangle, \quad (3.6)$$

$$\begin{aligned} h_{mn}^l &= -\frac{l(l+1)}{k-2} + \frac{(m+nk)^2}{k}, \\ \bar{h}_{mn}^l &= -\frac{l(l+1)}{k-2} + \frac{(m-nk)^2}{k}. \end{aligned} \quad (3.7)$$

In the following we are going to concentrate on those modes that satisfy the usual condition

$$(L_0 - \bar{L}_0) |T(r, \theta_L, \theta_R)\rangle = 0. \quad (3.8)$$

It is easy to see that this condition restricts the space of tachyon states to functions of the form

$$T(r, \theta_L, \theta_R) = T(r, \theta) + \tilde{T}(r, \tilde{\theta}), \quad (3.9)$$

where

$$\theta = \frac{1}{2}(\theta_L + \theta_R), \quad \tilde{\theta} = \frac{1}{2}(\theta_L - \theta_R). \quad (3.10)$$

We can thus introduce two separate tachyon fields $T(r, \theta)$ and $\tilde{T}(r, \tilde{\theta})$ that depend only on r and one θ -coordinate, and are two-dimensional scalar fields. The tachyon field $T(r, \theta)$ can be expanded in matrix elements of the continuous series, $T_m^\lambda(r, \theta) = \langle l, \frac{m}{2} | g(r, \theta, -\theta) | l, -\frac{m}{2} \rangle$ with $l = -\frac{1}{2} + i\lambda$. Similarly, the dual tachyon field $\tilde{T}(r, \tilde{\theta})$ is built out of the matrix elements $\tilde{T}_n^l(r, \tilde{\theta}) = \langle l, \frac{n}{2} k | g(r, \tilde{\theta}, \tilde{\theta}) | l, \frac{n}{2} k \rangle$. In the expansion of this tachyon field also discrete representations contribute. From this expansion in modes we deduce that the coordinate $\tilde{\theta}$ should be considered to have period $2\pi/k$.

Finally, notice that there is a clear geometrical distinction between the two θ variables on the group manifold $SL(2, R)$. The first coordinate θ is a truly angular variable which has period 2π in order to avoid a conical singularity at $r = 0$. On the other hand, $\tilde{\theta}$ parametrizes a non-contractable loop in $SL(2, R)$, and need not have a definite period when we allow ourselves to work on the universal covering group $\widetilde{SL}(2, R)$. The period $2\pi/k$ of $\tilde{\theta}$ is only indirectly determined by duality considerations (see later). It is possible to change the period of the Euclidean time by applying an orbifold construction. Since the CFT has a global $U(1)$ symmetry we can mod out by any Z_N subgroup, thereby introducing an orbifold singularity at $r = 0$. By repeating this construction one can build a theory for any rational multiple of the original compactification radius.

3.1. THE MOMENTUM MODES AND THEIR TARGET SPACE ACTION.

Let us take a closer look at the momentum states. In the conventional sigma-model approach a tachyon field is described by a target-space effective action of the form

$$S[T] = \int d^2x e^\Phi \sqrt{G} (G^{ij} \partial_i T \partial_j T - 2T^2), \quad (3.11)$$

where G_{ij} and Φ are the background metric and dilaton field. We now like to combine this point of view with the more abstract group theoretical discussion of the last section. The link between these two approaches is provided by the L_0 operator, which we would like to identify with the target space Laplacian of the sigma model. That is, we would like to have that

$$L_0 = -\frac{1}{2e^\Phi \sqrt{G}} \partial_i e^\Phi \sqrt{G} G^{ij} \partial_j, \quad (3.12)$$

The precise form of L_0 for the momentum modes follows from our discussion of the previous section. When acting on the tachyon field $T(r, \theta)$ the L_0 operator given in (3.4)

takes the simpler form

$$L_0 = -\frac{1}{k-2} \left[\frac{\partial^2}{\partial r^2} + \coth r \frac{\partial}{\partial r} + \left(\coth^2 \frac{r}{2} - \frac{2}{k} \right) \frac{\partial^2}{\partial \theta^2} \right] \quad (3.13)$$

We are now in a position to determine the effective metric and dilaton fields by comparing this result with (3.12). It is a simple exercise to determine the background that reproduces this Laplacian. The solution for G_{ij} and Φ is of the form

$$ds^2 = \frac{1}{2}(k-2) \left[dr^2 + \beta^2(r) d\theta^2 \right],$$

$$\Phi = \log(\sinh r / \beta(r)). \quad (3.14)$$

where $\beta(r)$ is given by

$$\beta(r) = 2 \left(\coth^2 \frac{r}{2} - \frac{2}{k} \right)^{-\frac{1}{2}} \quad (3.15)$$

In leading order in $1/k$ this coincides with the results (1.6) obtained in [5, 3]. It seems a reasonable conjecture that the above metric and dilaton field represent the exact solution to the sigma model β -function equations. This conjecture has been confirmed upto three loop level in [14]. Notice that in the critical case, $k = 9/4$, the $1/k$ corrections are substantial.

3.2. THE WINDING MODES AND DUALITY

From the conformal field theory point of view the momentum and winding vertex operators are treated on equal footing. They only differ in the relevant sign between the left and right quantum numbers ω_L and ω_R . The conformal dimensions and all other properties of the vertex operators are invariant if we change the sign of, say, ω_R . This is the general way in which target space duality, like the familiar $R \rightarrow 1/R$ symmetry of a compactified boson, is implemented in conformal field theory: the right-moving representations are replaced by their complex conjugates, whereas the left-movers are left invariant. In this particular case, the $SL(2, R)$ currents of the coset CFT transform under duality as

$$J^a \rightarrow J^a, \quad \bar{J}^a \rightarrow -\bar{J}^a, \quad (3.16)$$

which is clearly a symmetry of the theory. Thus, in terms of gauged WZW models, duality relates the two possible anomaly free gauge symmetries: the axial symmetry

$$g \rightarrow hgh, \quad (3.17)$$

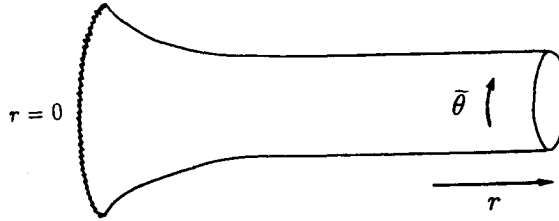


Fig. 4: The geometry in which the string winding modes propagate is related to that in fig. 3 by $R \rightarrow 1/R$ duality. In coordinates (r, θ) it has a curvature singularity, which for $k \rightarrow \infty$ occurs at $r = 0$.

and the vector symmetry

$$g \rightarrow hgh^{-1}. \quad (3.18)$$

However, when considered as sigma-models these two group actions lead to qualitatively different target spaces. The first action has no fixed points and therefore describes a regular target manifold. This is no longer true for the second group action, which has the identity element as a fixed point. Here we expect the manifold to develop a singularity. As conformal field theories these two models are completely equivalent, and thus that the two target space interpretations are equally valid.

The dual target space, corresponding to (3.18), in which the role of the momentum and winding modes are interchanged, is that in which the dual field $\tilde{T}(r, \tilde{\theta})$ propagates. For this dual tachyon the L_0 equation takes the form

$$L_0 = -\frac{1}{k-2} \left[\frac{\partial^2}{\partial r^2} + \coth r \frac{\partial}{\partial r} + \left(\tanh^2 \frac{r}{2} - \frac{2}{k} \right) \frac{\partial^2}{\partial \tilde{\theta}^2} \right] \quad (3.19)$$

Analogous to our discussion of the momentum modes, we can give an interpretation of this L_0 equation as a wave equation for a massless particle coupled to a specific G, Φ background. In this case the resulting geometry is again of the form (3.14), where we now find

$$\beta(r) = 2 \left(\tanh^2 \frac{r}{2} - \frac{2}{k} \right)^{-\frac{1}{2}}. \quad (3.20)$$

This metric and dilaton field now are singular at $r = \operatorname{arctanh} \sqrt{2/k}$. For large k we have

$$ds^2 = dr^2 + 4 \coth^2 \frac{r}{2} d\tilde{\theta}^2. \quad (3.21)$$

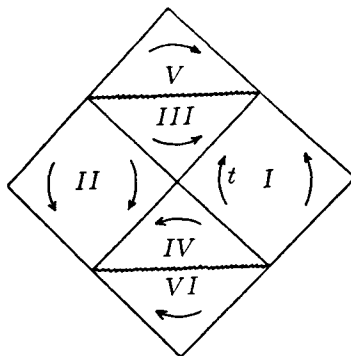


Fig. 5: Penrose diagram of the fully extended black hole, including the regions V and VI behind the singularity. The 'time' flow $\partial_t = u\partial_u + v\partial_v$ is indicated for all six regions.

The vortices are moving on a dual manifold that looks like a 'trumpet' (see fig. 3). In fact, in the $k \rightarrow \infty$ limit the two manifolds are related by the familiar $R \rightarrow 1/R$ duality transformation applied to the coordinate θ , but our analysis suggests that $1/k$ corrections modify this transformation. Another way to relate the two solutions, that is valid to all orders in $1/k$, is by the transformation $r \rightarrow r + i\pi/2$, $\theta \rightarrow \tilde{\theta}$.

3.3. SELF-DUALITY OF THE MINKOWSKI BLACK HOLE

It is natural to ask whether these stringy duality properties of the Euclidean have reflections in the Minkowskian theory. To address this question, let us now take serious the regions V and VI behind the singularities (see fig. 5). The metric in these regions describes a negative mass black hole with a naked singularity, and is in fact the analytic continuation of the trumpet which featured in the previous section. This leads to a remarkable interpretation of the duality transformation that interchanges the momentum and winding modes of the Euclidean theory. When analytically continued this transformation maps the Minkowskian black hole onto itself via the identification

$$y \rightarrow 1 - y, \quad t \rightarrow t, \quad (3.22)$$

where y is the radial coordinate $y = uv$. Under duality the regions I and V are interchanged, whereas region III is mapped onto itself. This interchanges in particular the

horizon ($y = 0$) with the singularity ($y = 1$). We note that region *III* can be seen as the analytic continuation of a metric of the form (for $k \rightarrow \infty$) $ds^2 = dr^2 + \tan^2 r d\theta^2$. As a CFT this describes a fractional level $SU(2)/U(1)$ coset, *i.e.* a parafermionic model, and with our definition of duality this model is ‘self-dual.’

The self-duality of the complete black hole string theory is most immediate using its formulation via the coset construction. As just explained, duality relates the group actions $g \rightarrow hgh$ and $g \rightarrow hgh^{-1}$, and both group action play a role in this model: one corresponds to the gauge symmetry, while the other represents the isometry generated by the Killing vector $\partial/\partial t$. For the non-compact choice of h both actions have fixed points, namely $SL(2, R)$ elements g with either $uv = 0$ or $ab = 1 - uv = 0$, where we made use of the parametrization (2.2). The fixed point of the gauge symmetry is the singularity, and the horizon is invariant under flow of the Killing vector. Self-duality is the statement that the actions $g \rightarrow hgh$ and $g \rightarrow hgh^{-1}$ are equivalent, after the transformation $y \rightarrow 1 - y$. This latter transformation can be seen to act on the group elements $g(z, \bar{z})$ as

$$g \rightarrow \epsilon \cdot g, \quad \epsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (3.23)$$

So, roughly speaking, the vertex operators in region *I* are the analytic continuations of the Euclidean *momentum* modes, whereas the operators in region *V* can be seen as analytic continuations of the Euclidean *winding* modes. We stress that, as it stands, duality does *not* give identifications between the modes in the physical region *I*, and region *V*. It simply expresses the equivalence of two, at first sight different, formulations of the black hole CFT.

4. QUANTIZATION OF THE EFFECTIVE TARGET SPACE THEORY

One of the interesting problems in black hole physics concerns the back reaction of the black hole geometry resulting from infalling or outgoing matter. We know that the mass of the black hole increases when it absorbs an infalling particle from infinity, while its mass decreases if it emits a Hawking particle. It would be extremely interesting if one could obtain an explicit description of this back reaction in string theory. Unfortunately, however, the present formalism of string theory is not well suited for describing such processes. Therefore, we could instead become less ambitious, and just consider the effective target space theory as an interesting toy model. Here one can perhaps obtain an explicit quantum mechanical description of the Hawking evaporation process, and furthermore address interesting questions in quantum gravity, such as the issue of quantum coherence or the physics near space-time singularities, albeit in a simplified context.

In the following we will make some remarks concerning the quantization of the effective target space action, which may become helpful in studying these questions. First we will concentrate of the gravitational sector of the model, containing only the metric and the dilaton field, described by the action*

$$S = \int d^2x e^{\Phi} \sqrt{G} (R + (\nabla\Phi)^2 - 1) \quad (4.1)$$

This action describes a topological model without local degrees of freedom. Moreover, the space of classical solutions to its equations of motion is just one-dimensional and consists of the black hole solutions

$$ds^2 = \frac{du dv}{M - uv}, \quad (4.2)$$

$$\Phi = \log(M - uv). \quad (4.3)$$

where the parameter M describes the mass of the black hole. One expects that the quantization of this theory should be rather simple, and that, as long as no second quantized matter is present, the classical black hole solutions remains stable in the quantum theory and correspond to static quantum states. As the next step, one of course would like to study explicitly what happens if we add matter to this system, since this will give rise to the Hawking effect and the black hole solution (4.2) will become quantum mechanically unstable.

*Here we used a field redefinition to put all constants in the action equal to 1.

We will discuss two quantization procedures. First we describe a canonical quantization method, which makes use of a first order formalism. Next we will make some comments about the conformal gauge formulation, which appears to be more suitable for including the back reaction due to second quantized matter.

4.1. CANONICAL QUANTIZATION

The following method of canonical quantization of the dilaton/gravity system (4.1) exploits the fact that the model is in fact rather similar to 2+1-dimensional gravity: indeed, the above action is almost identical to the dimensional reduction of the 2+1-dimensional Einstein action. As we will now show, the above theory is indeed exactly soluble in a similar way as 2+1-dimensional gravity was made manifestly soluble by Witten [16], namely by rewriting the model as a topological gauge theory of the relevant Poincaré group.

An ISO(1,1) gauge theory.

The Poincaré group in 1+1 dimensions is $ISO(1,1)$, and the corresponding gauge field formulation of the metric makes use of a zweibein $e^\pm$ and spin connection ω . Let us consider the following action

$$S = \int d^2x \left[X^0 d\omega + X^\mp D e^\pm - e^+ \wedge e^- \right] \quad (4.4)$$

Here the X^a are three scalar fields, playing the role of lagrange multipliers. In particular, fields $X^\pm$ impose via their equation of motion the two 'torsion constraints'

$$D e^\pm \equiv d e^\pm \mp \omega \wedge e^\pm = 0 \quad (4.5)$$

The above action describes a $ISO(1,1)$ gauge theory; the gauge transformations read

$$\begin{aligned} \delta\omega &= d\alpha & \delta X^0 &= \xi^+ X^- - \xi^- X^+ \\ \delta e^\pm &= D\xi^\pm \pm \alpha e^\pm & \delta X^\pm &= \xi^\pm \pm \alpha X^\pm \end{aligned} \quad (4.6)$$

and leave (4.4) invariant. We now claim that via the following identifications

$$\begin{aligned} X^0 &= e^\Phi \\ G &= \eta_{ab} \frac{e^a \otimes e^b}{X^0} \end{aligned} \quad (4.7)$$

the actions (4.1) and (4.4) are physically equivalent. To see this, note that the torsion constraints (4.5) can be used to solve for the spin connection ω in terms of the zweibein $e^\pm$. Inserting solution back into the gauge theory action (4.4), it can be rewritten in terms of original metric G and dilaton $\Phi = \log X^0$, and after an easy calculation one indeed recovers (4.1).

As a further check, let us determine the space of classical solutions to (4.4). The equations of motion of $e^\pm$ and ω are resp.

$$\begin{aligned} e^\pm &= \pm DX^\pm \\ dX^0 &= X^+ e^- - X^- e^+ \end{aligned} \quad (4.8)$$

These equations imply that the quantity M defined by

$$M = X^0 + X^+ X^- \quad (4.9)$$

is constant

$$dM = 0 \quad (4.10)$$

Hence the general solution for the original metric G is

$$G = \frac{DX^+ \otimes DX^-}{M - X^+ X^-} \quad (4.11)$$

Comparing this with the black hole solution (1.3) we find that the fields $X^\pm$ and the light-cone coordinates u, v used in the previous section are related via

$$\begin{aligned} u(x) &= X^+(x) e^{-\int^x \omega} \\ v(x) &= X^-(x) e^{\int^x \omega} \end{aligned}$$

The interpretation of the parameter M as the mass of the black hole will become more clear in a moment.

Physical States

Canonical quantization of the $ISO(1, 1)$ gauge theory is performed most conveniently in the gauge

$$e_t^\pm = 1 \quad \omega_t = 0 \quad (4.12)$$

To this end, let us rewrite the action (4.4) in the Hamiltonian form

$$S = \int dt dx \left[X^0 \partial_t \omega_x + X^\mp \partial_t e_x^\pm + \omega_t \mathcal{G}^0 + e_t^\pm \mathcal{G}^\mp \right] \quad (4.13)$$

Here the time-components of the gauge field are lagrange multipliers imposing the constraints

$$\begin{aligned} \mathcal{G}^0 &= \partial_x X^0 + e_x^+ X^- - e_x^- X^+ \\ \mathcal{G}^\pm &= D_x X^\pm \mp e_x^\pm \end{aligned} \quad (4.14)$$

Before imposing these constraints, the phase space of the theory consists of the space components of the three gauge fields and the three X -fields. From (4.13) we immediately read of that they are each others canonical conjugates

$$\begin{aligned} [e_x^\pm(x), X^\mp(y)] &= i\delta(x-y) \\ [\omega_x(x), X^0(y)] &= i\delta(x-y) \end{aligned} \quad (4.15)$$

Physical states are functionals of a maximal commuting subset of these variables and are further selected to satisfy the constraints (4.16)

$$\mathcal{G}^a \Psi_{phys} = 0 \quad (4.16)$$

where $\mathcal{G}^a$ are quantum operators. If we choose the physical states to be functionals of the X^a field variables, the general solution to (4.16) reads

$$\Psi[X^a] = e^{\int X^+ \partial_x X^-} \delta(\partial_x M) \Psi_{red}(M) \quad (4.17)$$

where $M = X^0 + X^+ X^-$. Thus physical states are characterized by a function $\Psi_{red}(M)$ of the mass of the black hole. Hence we indeed find what we expected from the classical picture, namely that for each value for the mass M of the black hole there exists a corresponding physical state $|M\rangle$.

Interaction with Point-Particles.

It is interesting to study the back reaction of the quantum black hole geometry due to infalling matter. The coupling to first quantized point particles is naturally incorporated in the gauge theory formulation via the inclusion of Wilson line operators. The operator

$$W(x) = \exp\left(\int_{-\infty}^x e^+ p^- + e^- p^+ + \omega J\right) \quad (4.18)$$

represents the interaction with a particle coming in from infinity, that has arrived at the position x . To analyze the effect of this particle on the black hole geometry, let us for definiteness consider a massless left moving particle. Its worldline action reads

$$S[p^-, q^+] = \int dt [p^- \partial_t q^+ + e^+ p^- + \omega p^- q^+] \quad (4.19)$$

From the canonical commutators (4.15) we see that this point-particle creates a discontinuity in one of the ‘coordinates’ X^- and the mass M of the black hole. In passing through the world-line of the point particle, the following shifts occur

$$\begin{aligned} X^- &\rightarrow X^- + p^- \\ M &\rightarrow M + p^- q^+ \end{aligned} \quad (4.20)$$

Hence we see that the back-reaction of the black hole geometry due to the stress-energy of a point-particle is two-fold. On the one hand, the position of the black hole is shifted with an amount proportional to the $+$ -momentum p^+ of the particle. In addition, the black hole mass M increases with an amount $\Delta E = p^+ q^-$, which is indeed the energy of the incoming particle when measured in the asymptotically flat region. In a moment we will rederive this result via a more conventional calculation, by solving the classical field equations of the original theory in presence of an infalling particle.

Coupling to Second Quantized Matter

There are several ways of coupling second quantized matter to the dilaton/gravity system, while preserving the $ISO(1,1)$ gauge symmetry. One way is to represent the matter degrees of freedom via two conjugate fields, a one-form $\pi(x, q)$ and a scalar $\phi(x, q)$, which both in addition depend on two internal coordinates q^a . As the action for these matter fields one then takes

$$S_m[\pi, \phi] = \int \pi(x, q) \wedge [d + e \cdot p + \omega(q \times p)] \phi(x, q) \quad (4.21)$$

where the integral runs over both x^a and d is the x -exterior derivative and

$$p_a = \frac{d}{dq^a}. \quad (4.22)$$

The internal space parametrized by the coordinate q serves as a representation space for the $ISO(1,1)$ transformations. In the above matter action the field π plays a dual role:

it is the canonical momentum of φ as well as a lagrange multiplier for the ‘soldering constraint’

$$[d + e \cdot p + \omega(q \times p)]\varphi(x, q) = 0, \quad (4.23)$$

which relates the q and the x -dependence of $\varphi(x, q)$. This way of coupling matter clearly does not correspond to including the tachyon field T in the action. However, we expect it to be equivalent to adding a conformal scalar field to the dilaton/graviton system.

As remarked before, one would expect that when quantizing the theory with matter one would be able to find Hawking radiation. The canonical quantization of the above theory is still rather straightforward. Following the same steps as before, one can construct solutions to the (modified) physical constraints. We omit this analysis here, mainly because the result is somewhat disappointing: one does not find any trace of Hawking radiation or of an instability of the black hole solution. Apparently, via the above quantization method one implicitly uses a regularization in which matter does not produce Hawking radiation. To understand better what is going on here, we will return to the original dilaton/graviton system and consider it in the conformal gauge.

4.2. CONFORMAL GAUGE QUANTIZATION

We will now make some remarks on how the above results are related to the more conventional conformal gauge treatment. The conformal gauge is particularly convenient for discussing the Hawking phenomenon, since it is well known to be intimately connected with the conformal anomaly [17].

Geometry of Infalling Massless Particle

The theory (4.1) was first analyzed in the conformal gauge in [4]. In this gauge the action takes the form

$$S_0[\sigma, \Phi] = \frac{1}{2\pi} \int dudv e^{\Phi} \partial_u \partial_v (\sigma + \Phi) + e^{\sigma + \Phi} \quad (4.24)$$

where σ is the conformal factor:

$$G = e^{\sigma} dudv.$$

The equations of motion of this action are conveniently written as

$$\begin{aligned} \partial_v \partial_u (\Phi + \sigma) &= 0 \\ \partial_v \partial_u e^{\Phi} + e^{\Phi + \sigma} &= 0 \end{aligned} \quad (4.25)$$

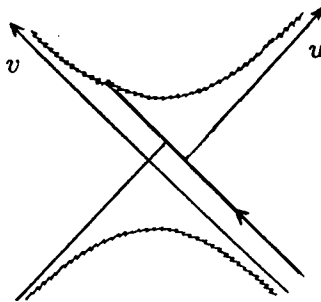


Fig. 6: The geometry of an infalling massless particle, described by the metric $ds^2 = e^\sigma du dv$ given in eqn (4.27).

In addition there is the physical constraint that the total stress-tensor vanishes. This leads uniquely to the black hole solution of [4, 3].

It is instructive to reinvestigate in this language the effect on this geometry due to a single massless infalling particle. In this case the stress-energy constraint reads

$$e^\Phi (\partial_u \sigma \partial_u \Phi - \partial_u^2 \Phi) = p_- \delta(u - q_+) \quad (4.26)$$

It is readily checked that the general solution to these equations is given by

$$e^\Phi = e^{-\sigma} = \begin{cases} M - uv & u < q_+ \\ (M + p_- q_+) - (v + p_-)u & u > q_+ \end{cases} \quad (4.27)$$

This geometry, depicted in fig. 6, consists of two vacuum solutions glued together along the world-line of the particles. Note that the patching condition coincides with that found in the previous section using the topological $ISO(1,1)$ gauge theory formulation. In interesting limiting case of the above solution is obtained by putting $M = 0$, where it describes the formation of a black hole due to an infalling particle.

Quantum Theory: Is the black hole stable or not?

In the quantum theory we have to account for the path integral measure. It turns out that there is some freedom of choice here, and correspondingly there appear to be two

inequivalent quantization procedures. Namely, as usual in the conformal gauge we need to include a (b, c) ghost system and the trouble comes from the conformal anomaly of this ghost partition function.

As a consequence of the conformal anomaly, the quantum effective action is no longer given by (4.24), but includes an additional Liouville term. There are now two reasonable choices for which conformal factor one uses to define the measure, namely, besides the one defined with $G = e^\sigma du dv$, one could also use the measure defined with the metric $\tilde{G} = e^\Phi G$. With the first choice, the quantum action becomes (here $c = -26$).

$$S_1 = S_0[\sigma, \Phi] + \frac{c}{24\pi} \int \partial_u \sigma \partial_v \sigma \quad (4.28)$$

while with the second definition of the path-integral measure one gets

$$S_2 = S_0[\sigma, \Phi] + \frac{c}{24\pi} \int \partial_u (\sigma + \Phi) \partial_v (\sigma + \Phi) \quad (4.29)$$

The crucial qualitative difference between these two actions is that, while the black hole geometry (4.2) is still a classical solution of the second action $S_2^\dagger$, this is no longer true for the first action S_1 .

The modification of the black hole solution of S_1 has an (un)physical interpretation in terms of ‘Hawking radiation of ghosts.’ At first the occurrence of such a phenomenon may perhaps sound rather unacceptable, but it is clear that the same reasoning [15] that demonstrates the existence of black hole radiation for ordinary matter fields also applies for ghost fields. Moreover, although gauge invariance forbids us to directly detect ghost-particles, they do still contribute to the stress-energy. What is further unusual, however, is that the sign of the conformal anomaly of the ghosts is opposite to that of normal matter, and thus in the pure dilaton gravity model the black hole mass in fact *increases* in time due to the thermal emission of the ghost particles. Hence we conclude that, whereas it seemed perhaps more reasonable at first to define the functional measure $[dG d\Phi]$ using the standard metric G , the alternative choice using $\tilde{G} = e^\Phi G$, leading to the second effective action, seems to lead to physically more acceptable conclusions. It should also be clear that the canonical quantization method described above is equivalent to the second procedure and not to the first.

[†]Note that in the black hole solution $\sigma + \Phi = 0$, which also minimizes the second term in S_2 .

Some Conclusions

Regardless of which of the two quantization methods we adopt, we can of course add to the metric/dilaton theory a matter system. This will give rise to black hole radiation, provided we define the path integral measure with the help of the un-rescaled metric G .[†] This Hawking effect can be studied most explicitly if we take the matter theory to be conformally invariant. For example we can add to the action a bunch of conformal scalars

$$S^{(m)} = \sum_{i=1}^N \int \partial_u x_i \partial_v x_i \quad (4.30)$$

Their presence will produce, via the conformal anomaly, an additional term proportional to the Liouville action of σ to the total effective action of the metric, or equivalently, a contribution to the stress-energy of the form

$$T_{uu}^{(m)} = T(u) - \frac{N}{12} (\frac{1}{2} \partial_u \sigma \partial_u \sigma + \partial_u^2 \sigma) \quad (4.31)$$

and similar for $T_{vv}^{(m)}$, while $T_{uv}^{(m)} = \frac{N}{12} \partial_u \partial_v \sigma$. In (4.31) the analytic part $T(u)$ is determined by the initial conditions on the x -fields, *i.e.* their in-state. For example, in the far past we may choose the vacuum boundary condition, $T_{uu}^{(m)} \rightarrow 0$ on $\mathcal{I}^-$. It is now well-known that, if we consider for example the geometry in fig. 6, then, if we neglect the back reaction on the geometry, this vacuum state will evolve into a final situation where on $\mathcal{I}^+$ there will be a non-vanishing expectation value for the stress-tensor T [15, 17]. The corresponding modification of the equation of motion for the gravitational field and the dilaton is easy to write down, but, unfortunately, hard to solve explicitly. Clearly it would be interesting to explicitly find this solution, which would describe the geometry of an evaporating black hole. There are ways to try and solve it perturbatively, either in $N/12$ or in $\frac{1}{N}$. In both limits it can be seen that, while initially the infalling particle increases the black hole mass by an amount $\delta E = p^- q^+$, there is an out-flux of energy due to Hawking radiation which at the end will result in a final situation where all mass has been emitted to $\mathcal{I}^+$ and the black hole has disappeared.

Aspects of this process have been studied recently in the large N limit in [18]. The central observation of these authors is that the theory in fact has two qualitatively different

[†]Indeed, the explanation for the fact that in the $ISO(1,1)$ gauge theory formalism second quantized matter does not produce any Hawking radiation is that the corresponding canonical quantisation method implicitly uses for the matter functional measure the one defined with the flat metric $\hat{G} = \eta_{ab} e^a \otimes e^b$.

asymptotic regions. One is the region far away from the black hole, where the dilaton ϕ approaches $+\infty$ and the classical term S_0 of the action dominates. In the other asymptotic region near to the black hole the dilaton field tends to $-\infty$; here the theory is still well-behaved and in fact dominated by the induced Liouville effective action. Based on this observation, in [18] a picture of the black hole ‘formation and evaporation’ process is proposed, in which infalling matter particles already lose all their mass due to the emission of Hawking radiation before the black hole has been formed. It is further argued that this process should be describable by a unitary S -matrix.

In conclusion, we have seen that the two-dimensional dilaton gravity theory without matter is exactly soluble. When coupled to conformal matter it is still tractable in certain limits, thereby providing an interesting toy model for studying some fundamental questions in black hole physics.

Acknowledgements

The results described in the first part of these notes have been obtained in collaboration with R. Dijkgraaf and E. Verlinde. It is further a pleasure to acknowledge fruitful discussions with C. Callan, J. Distler, G. Horowitz, D. Kutasov, I. Klebanov, M. Roček, A. Strominger, N. Seiberg, S. Shenker and E. Witten. This research is supported by NSF Grant PHY90-21984.

References

- [1] For a review of the recent developments and references, see the lectures by Igor Klebanov in this volume.
- [2] C. G. Callan, D. Friedan, E. J. Martinec and M. J. Perry, Nucl. Phys **B 262** (1985) 593.
- [3] E. Witten, Phys.Rev **D44** (1991) 314.
- [4] G. Mandal, A.M. Sengupta, and S.R. Wadia, Mod. Phys. Lett. **A6** (1991) 1685
- [5] S. Elitzur, A. Forge, and E. Rabinovici, Nucl. Phys. **B 359** (1991) 581.
- [6] R. Dijkgraaf, E. and H. Verlinde, ‘String Propagation in a Black Hole Geometry,’ preprint PUPT-1252

- [7] K. Gawedzki and A. Kupiainen, Nucl. Phys **B 320** (1989) 625; K. Gawedski, preprint, october 1991
- [8] P. Goddard, A. Kent, and D. Olive, Comm. Math. Phys. **103** (1986) 105.
- [9] L. Dixon, J. Lykken, and M. Peskin, Nucl. Phys. **B325** (1989) 329; I. Bars and D. Nemeschansky, preprint IASSNS-HEP-90/15 (1990); I. Bars, Nucl. Phys **B334** (1990) 125.
- [10] J. Distler and P. Nelson, Princeton preprint (1991)
- [11] N.J. Vilenkin, 'Special Functions and the Theory of Group Representations,' (AMS, 1968).
- [12] T. H. Buscher, Phys. Lett **B201** (1988) 466; A.A. Tseytlin, 'Duality and Dilaton,' John Hopkins preprint JHU-TIPAC-91008.
- [13] A. Giveon, 'Target Space Duality and Stringy Black Holes,' Berkeley preprint LBL-30671 (1991).
- [14] A. Tseytlin, Phys. Lett. **B 268** (1991) 175
- [15] S.W. Hawking, Comm. Math. Phys. **43** 199 (1975) 199; J.B. Hartle and S.W. Hawking, Phys. Rev. **D 13** (1976) 2188.
- [16] E. Witten, Nucl. Phys. **B 311** (1988) 46
- [17] S. Christensen and S. Fulling, Phys. Rev **D15** (1977) 2088;
see also N. Birrel and P. Davies, 'Quantum Fields in Curved Space' (Cambridge University Press, 1982)
- [18] C. Callan, S. Giddings, J. Harvey and A. Strominger, 'Evanescent Black Holes,' preprint UCSB-TH-91-54

Supersymmetric String Solitons

Curtis G. Callan, Jr.

*Department of Physics, Princeton University
Princeton, NJ 08544*

Internet: cgc@pupphy.princeton.edu

Jeffrey A. Harvey

*Enrico Fermi Institute, University of Chicago
5640 Ellis Avenue, Chicago, IL 60637*

Internet: harvey@curie.uchicago.edu

Andrew Strominger

*Department of Physics, University of California
Santa Barbara, CA 93106*

Bitnet: andy@voodoo

Abstract

The subject of these lectures is the construction of supersymmetric soliton solutions to superstring theory. A brief review of solitons and instantons in supersymmetric theories is presented. Yang-Mills instantons are then used to construct soliton solutions to heterotic string theory of various types. The structure of these solutions is discussed using low-energy field theory, sigma-model arguments, and in one case an exact construction of the underlying superconformal field theory.

1. Introduction

The theme of most of the lectures given at this school has been the study of simplified or “toy” models of string theory. The motivation of course is to better understand what a fundamental and non-perturbative formulation of string theory should look like through the use of simple models. These lectures describe a different, but hopefully complementary, approach to the study of non-perturbative string theory through the development of semi-classical techniques.

There are several reasons why this is an interesting enterprise. First, in realistic unified string theories there is little hope that we will obtain exact non-perturbative results. As in most field theories, probably the best we can hope for is to obtain approximate non-perturbative results through the use of semi-classical techniques. Still, even achieving this limited goal is bound to teach us many interesting things about the structure of string theory, as it has in field theory. Second, there are a number of fascinating technical issues in conformal field theory which must be resolved before this program can be carried out. These include the proper treatment of collective coordinates and the proper definition of the mass or action in string theory. Related issues may also arise in the treatment of back reaction in the exact black hole solution [1] discussed in the lectures of H. Verlinde at this school. Third, the solitons are in many cases the endpoints of black hole (or black p -brane) evaporation, and they are therefore an essential ingredient in resolving the fascinating issues surrounding Hawking radiation and coherence loss in string theory. Finally, we may learn something dramatically new about string theory through the study of semi-classical solutions. For example, the conjecture [2] that strongly coupled string theory is dual to weakly coupled fivebranes would have fascinating consequences if true.

The study of semi-classical string theory is still in its infancy. We do not yet know how to classify semi-classical solutions directly through stringy topological invariants nor do we have all the techniques necessary for extracting physical quantities by expanding string theory about these solutions. What has been accomplished recently is to find solutions to string theory, some of which can be studied directly as (super) conformal field theories, which at large distances are solutions to the low-energy effective string field theory equations of motion and which have the properties of solitons and instantons. One goal of these lectures is to discuss these solutions and to indicate how they might be used to study some of the issues raised above.

In order to make these lectures reasonably self-contained we will start in Section 2 by reviewing certain features of solitons in field theory that will be important in what follows. Section 3 reviews aspects of Yang-Mills instantons in supersymmetric theories. The Montonen-Olive conjecture of a weak-strong coupling duality in $N = 4$ Yang-Mills is described in Section 4, along with comments on its possible relevance to string theory. Section 5 discusses solitons and instantons in low-energy heterotic string theory and the special role played by fivebrane solitons. The geometry and charges of these solutions is explained in Section 6. Symmetries, non-renormalization theorems, and other aspects of the underlying worldsheet sigma model are discussed in Section 7. An algebraic construction of certain symmetric solutions is given in Section 8, and the appearance of an exotic $N = 4$ algebra is described. Section 9 briefly reviews some additional supersymmetric solitons. Concluding comments are made in Section 10.

2. Soliton Review

Roughly speaking, solitons are static solutions of classical field equations in D space-time dimensions which are localized in $(D - 1) - d$ spatial coordinates and independent of the d other spatial coordinates. The usual case is $d = 0$ and the soliton then has many characteristics of a point particle. For arbitrary d the solution is called a d -brane and describes an extended object. In four spacetime dimensions $d = 1$ corresponds to a string and $d = 2$ to a membrane or domain wall. Such objects are of interest in various cosmological scenarios.

Solitons are usually characterized by the following properties [3]. First, they are non-perturbative. They are solutions to non-linear field equations which cannot be found by perturbation of the linearized field equations. In addition, their mass (or mass per unit d -volume) is inversely proportional to some power of a dimensionless coupling constant. As a result, they become arbitrarily massive compared to the perturbative spectrum at weak coupling. Thus quantum effects due to exchange of solitons will be non-perturbative effects which vanish to all orders in perturbation theory. Second, solitons are characterized by a topological rather than a Noether charge. Finally, soliton solutions typically depend on a finite number of parameters called moduli which act as coordinates on the moduli space of soliton solutions of fixed topological charge.

A simple and standard example which illustrates these features is the “kink” solution in $D = 1 + 1$ spacetime dimensions. Start with the Lagrangian

$$\mathcal{L} = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - U(\phi) \quad (2.1)$$

with the potential given by $U(\phi) = \lambda(\phi^2 - m^2/\lambda)^2/4$. The dimensionless coupling in this example is $g \equiv \lambda/m^2$. The conserved topological charge is

$$Q = \int_{-\infty}^{+\infty} dx j_0 \quad (2.2)$$

where $j_\mu = (\sqrt{g}/2)\epsilon_{\mu\nu}\partial^\nu\phi$ is the conserved topological current. Thus

$$Q = \frac{\sqrt{g}}{2}(\phi(+\infty) - \phi(-\infty)) \quad (2.3)$$

and is ± 1 for a kink (anti-kink) in which ϕ varies from the minimum of U at $\phi = \mp 1/\sqrt{g}$ at $x = +\infty$ to the minimum at $\phi = \pm 1/\sqrt{g}$ at $x = -\infty$.

To find the explicit form of the solution we consider the classical equation of motion for a static configuration

$$\phi'' = \frac{\partial U}{\partial \phi}. \quad (2.4)$$

For a solution with U and ϕ' vanishing at infinity we can integrate this once to find

$$\frac{1}{2}(\phi')^2 = U(\phi). \quad (2.5)$$

Integrating this equation with the previous choice of U yields the kink (anti-kink) solution

$$\phi_{K(K)}(x) = \pm \frac{m}{\sqrt{\lambda}} \tanh[m(x - x_0)/\sqrt{2}]. \quad (2.6)$$

The energy (rest mass) of this configuration is

$$E = \int dx \frac{1}{2}(\phi')^2 + U(\phi) = \frac{2\sqrt{2}}{3} \frac{m}{g} \quad (2.7)$$

so that the kink mass divided by the mass of the elementary scalar is proportional to $1/g$ showing the non-perturbative nature of the solution.

Also, in accordance with the earlier discussion, the solution for fixed Q and mass depends on a parameter x_0 , the center of mass coordinate of the soliton. In this example the existence of x_0 follows from translational invariance of the underlying field theory. The presence of x_0 introduces an important problem into the quantization of the theory expanded about ϕ_K . This is due to the fact that the quadratic fluctuation operator

$$\mathcal{O}_2 = \frac{\delta^2 S}{\delta \phi^2} |_{\phi=\phi_K} = \partial^2 + m^2 - 3\lambda \phi_K^2 \quad (2.8)$$

has a zero mode η_0 given by an infinitesimal translation of ϕ_K , $\eta_0 = \partial \phi_K / \partial x_0$. We can think of η_0 as a tangent vector to the curve in configuration space given by translation of ϕ_K . A zero mode of $\mathcal{O}_2$ will lead to a divergence in perturbation theory about ϕ_K of the form $\det^{-1/2} \mathcal{O}_2 = \infty$.

The solution is to separate out the dependence on x_0 by a change of coordinates in field space through the introduction of a collective coordinate. Naively we would expand ϕ in the kink sector as

$$\phi(x, t) = \phi_K(x - x_0, t) + \sum_{n=0}^{\infty} c_n(t) \eta_n(x - x_0) \quad (2.9)$$

where the η_n are a complete set of eigenfunctions of $\mathcal{O}_2$ and the $c_n(t)$ are time dependent coefficients which act as coordinates in configuration space. The collective coordinate method involves a change of coordinates from the $\{c_n, n = 0 \dots \infty\}$ to $\{x_0(t), c_n(t), n = 1 \dots \infty\}$ which promotes the modulus x_0 to a time-dependent coordinate. The expansion of ϕ is then given by

$$\phi(x, t) = \phi_K(x - x_0(t)) + \sum_{n=1}^{\infty} c_n(t) \eta_n(x - x_0(t)). \quad (2.10)$$

It is then possible to separate out explicitly the dependence on $x_0(t)$ and to quantize this as the center of mass coordinate of the soliton and to be left with a well defined perturbation theory for the non-zero modes.

In the more complicated examples we consider later there will be moduli resulting from symmetries of the underlying theory (translation invariance, scale symmetry, gauge symmetry, supersymmetry) as well as moduli which follow from index theorems but are not required by the underlying symmetries.

Two trivial modifications of this theory lead to solutions with somewhat different interpretations. We can add some spatial dimensions, say two, to obtain a static solution to 3 + 1 dimensional field equations which is independent of two of the coordinates, $\phi(x_1, x_2, x_3, t) = \phi_K(x_3)$. Such a solution describes a two-brane or what would be called in cosmology a domain wall. Alternatively we can remove the time dimension and reinterpret the solution $\phi_K(x)$ as a solution in Euclidean time of a 0 + 1-dimensional theory (i.e. quantum mechanics) $\phi_K(x) \sim x(t_E)$. The solution is then an instanton describing tunneling between two degenerate minima.

A far less trivial modification is to introduce fermions into the theory. Since we are interested in solitons in supersymmetric string theories, we introduce a supersymmetric coupling to fermions by taking the Lagrangian to be

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu \phi)^2 + \frac{1}{2}\bar{\psi}i\gamma^\mu \partial_\mu \psi - \frac{1}{2}V^2(\phi) - \frac{1}{2}V'(\phi)\bar{\psi}\psi \quad (2.11)$$

where ψ is a Majorana fermion and V is a function chosen so that the potential V^2 has two degenerate minima. For concreteness we may take $V = \lambda(\phi^2 - a^2)$. This theory has two chiral supercharges given by

$$Q_\pm = \int dx (\dot{\phi} \pm \phi')\psi_\pm \mp V(\phi)\psi_\mp \quad (2.12)$$

where $\psi_\pm$ are the left- and right-handed components of ψ .

It was shown in [4] that a careful treatment of boundary terms leads to the supersymmetry algebra

$$Q_+^2 = P_+, \quad Q_-^2 = P_-, \quad \{Q_+, Q_-\} = T \quad (2.13)$$

with $P_\pm = P_0 \pm P_1$, and where the central term T is equivalent to the topological charge defined earlier, although its precise functional form is different. The existence of this central term in the supersymmetry algebra in the soliton sector has important consequences. First, the relation

$$P_+ + P_- = (Q_+ + Q_-)^2 - T = (Q_+ - Q_-)^2 + T \quad (2.14)$$

implies a Bogomolnyi bound, $M \geq T/2$, where M is the rest mass. This bound is saturated precisely for those states $|s\rangle$ for which $(Q_+ \pm Q_-)|s\rangle = 0$, i.e. for states annihilated by some combination of the supersymmetry charges. Classically one can check that this does hold for the kink solution using (2.12) and (2.5). In fact, (2.5), or rather its square root could have been derived by demanding that the kink solution preserve a linear combination of supersymmetries. The fact that demanding unbroken supersymmetry leads to a sort of "square root" of the equations of motion will be put to great use later. Also, the fact that the kink solution is annihilated by one combination of the supercharges implies saturation of the Bogomolnyi bound. The other combination of supercharges does not annihilate

the kink state but instead produces a fermion zero mode in the kink background. This is because a supersymmetry variation of a solution to the bosonic equations of motion if it is nonzero produces a solution to the fermionic equations of motion in the bosonic background.

This simple example provides a paradigm for more complicated examples. In most known examples of solitons in supersymmetric theories one finds that half of the supercharges annihilate the classical solution leading to saturation of a Bogomolnyi bound, while the other half acting on the soliton produce fermion zero modes in the soliton background¹. In addition, searching for configurations which preserve some of the supercharges provides a shortcut to solving the full equations of motion since the resulting equations are typically first order as compared to the second order equations of motion.

In the kink example there is a single fermion zero mode. It also has interesting consequences for the spectrum of the theory. The mode expansion of the fermion field in the soliton sector will have the form

$$\psi = b_0 f_0 + \sum_{n \neq 0} b_n f_n \quad (2.15)$$

where f_0 is the zero mode. Canonical quantization implies that b_0 obeys $\{b_0, b_0\} = 1$ which in turn implies that the vacuum must consist of two degenerate states, $|+\rangle$ and $|-\rangle$ with $|\pm\rangle = b_0 |\mp\rangle$. We can think of the bosonic and fermionic zero modes (x_0, b_0) as providing coordinates on the supermoduli space of zero energy perturbations of the soliton.

3. Yang-Mills Instantons

After this brief review of solitons we turn our attention to instantons in Yang-Mills theory. These will provide the starting point for the construction of soliton solutions to string theory in ten dimensions, so we will quickly review some of the features of instanton solutions which are important in the later constructions.

We start with the Euclidean Yang-Mills action

$$S_E = \frac{1}{2g^2} \int d^4x \text{Tr} F_{\mu\nu} F^{\mu\nu} \quad (3.1)$$

and look for solutions of the equations of motion which approach pure gauge at infinity,

$$A_\mu \rightarrow g^{-1} \partial_\mu g. \quad (3.2)$$

Such configurations are labelled by the winding number of the map from S^3 into the gauge group G provided by the gauge function g of (3.2) evaluated at infinity. For simplicity we

¹ Although this is not always true. A fascinating counterexample [5] will be discussed in Section 9.

take $G = SU(2)$. It is simple to show that the action obeys a Bogomolnyi bound in this theory, $S_E \geq 8\pi^2 k/g^2$ where

$$k = \frac{1}{16\pi^2} \int d^4x \text{Tr} F_{\mu\nu} \tilde{F}^{\mu\nu} \quad (3.3)$$

is the winding number and $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\lambda\rho} F_{\lambda\rho}$. The bound is saturated for (anti) self-dual gauge fields obeying

$$F_{\mu\nu} = \pm \tilde{F}_{\mu\nu}. \quad (3.4)$$

The interpretation of these solutions is that they correspond to tunneling between degenerate minima labelled by elements of $\pi_3(SU(2))$. Applications of Yang-Mills instantons are described in [6] and [3].

To study the bosonic zero modes (moduli) of a solution to the self-dual equations we consider a fixed solution A_μ^0 of (3.4) and look for perturbations δA_μ which preserve (3.4) with a plus sign. Writing $A_\mu = A_\mu^0 + \delta A_\mu$ gives $F_{\mu\nu} = F_{\mu\nu}^0 + D_\mu \delta A_\nu - D_\nu \delta A_\mu$ where the covariant derivative D is evaluated using A^0 . Equation (3.4) then requires that

$$D_\mu \delta A_\nu - D_\nu \delta A_\mu - \epsilon_{\mu\nu\lambda\rho} D^\lambda \delta A^\rho = 0. \quad (3.5)$$

We are interested in solutions of (3.5) which are not of the form $\delta A_\mu = D_\mu \Lambda$, that is not pure gauge. The number of such solutions can be determined through the use of the Atiyah-Singer index theorem as discussed in [7]. Briefly, one considers the following elliptic complex

$$0 \xrightarrow{D^0} V^0 \xrightarrow{D^1} V^1 \xrightarrow{D_+^2} V_+^2 \xrightarrow{D^3} 0 \quad (3.6)$$

where V^0 is the space of zero-forms (scalars), V^1 is the space of one-forms (deformations of the instanton), and V_+^2 is the space of self-dual two-forms. D^0 and D^3 are the identity map and the projection map onto anti-self dual forms respectively. $D^1 = d + A^0$ and D_+^2 is the projection onto the self-dual part of the covariant exterior derivative (as in (3.5)). It is easy to check that $D_+^2 D^1 = 0$ iff $F = \tilde{F}$. We then define spaces

$$H^i = \frac{\text{Ker} D^{i+1}}{\text{Im} D^i} \quad (3.7)$$

and set $h^i = \dim H^i$. H^1 is the moduli space we are after, that is the space of solutions of (3.5) modulo solutions which are pure gauge. The Atiyah-Singer index theorem then gives the alternating sum of the h^i in terms of topological invariants. For $G = SU(2)$ one has $h^0 - h^1 + h^2 = -8k + 3$ for base manifold S^4 while for R^4 this sum is simply $8k$. For $G = SU(2)$ in these cases it is easy to show that $h^0 = 0$, that is there are no non-trivial solutions of $D_\mu \Phi = 0$ (For other groups and/or manifolds this is not necessarily true. Gauge connections for which $h^0 \neq 0$ are called reducible connections.) It is also possible to show using a simple positivity argument that $h^2 = 0$ so that on R^4 the dimension of the moduli space is just $8k$. In the later constructions it will be important that we are working on R^4 and not S^4 and that the total number of moduli is a multiple of four.

The construction of an Ansatz which contains all $8k$ parameters is a difficult and subtle mathematical problem whose solution can be found in [8]. For our purposes this will not be necessary. There is however a construction which yields $5k$ of the parameters and which plays an important role in the later constructions.

This ansatz involves writing the gauge field in terms of the derivative of a scalar field f as

$$A_\mu = \bar{\Sigma}_\mu^\nu \partial_\nu \ln f \quad (3.8)$$

where the $SU(2)$ indices have been suppressed and with $\bar{\Sigma}_{\mu\nu}$ a two by two anti-symmetric matrix which is anti-self-dual on the $\mu\nu$ indices. A simple construction of $\bar{\Sigma}$ which will be used later arises from embedding the $SU(2)$ gauge group in $SO(4)$. Letting $m, n = 1 \cdots 4$ be $SO(4)$ indices we have

$$\bar{\Sigma}_{\mu\nu}^{mn} = \frac{1}{2}(\delta_{\mu\nu}^{mn} - \frac{1}{2}\epsilon_{\mu\nu}^{mn}). \quad (3.9)$$

which is anti-self-dual both on the μ, ν indices and on m, n . This latter fact shows that the $\bar{\Sigma}$ lie in a $SU(2)$ subgroup of $SO(4)$.

Substituting this ansatz into the equation (3.4) results in a simple equation for f

$$\frac{1}{f} \square f = 0. \quad (3.10)$$

If f is non-singular then it is constant and hence $A_\mu = 0$. On the other hand, f can be singular and still solve (3.10) everywhere if the singularities in $\square f$ are cancelled by the $1/f$ factor. The general solution of this form is

$$f(x) = 1 + \sum_{i=1}^k \frac{\rho_i^2}{(x - x_i^0)^2} \quad (3.11)$$

and depends on $5k$ parameters $(x_{i\mu}^0, \rho_i)$ which give the location and the scale size of the instanton. The $3k$ parameters which are missing correspond roughly to the relative $SU(2)$ orientations of the k instantons plus an overall global $SU(2)$ orientation.

For $k = 1$ the solution (3.11) yields the gauge field

$$A_\mu = -2\rho^2 \bar{\Sigma}_{\mu\nu} \frac{(x - x_0)^\nu}{(x - x_0)^2((x - x_0)^2 + \rho^2)}. \quad (3.12)$$

The gauge field (3.12) is singular as $x \rightarrow x^0$ and does not approach pure gauge at infinity. However, by making a large gauge transformation one obtains a gauge field

$$A'_\mu = -2\Sigma_{\mu\nu} \frac{(x - x_0)^\nu}{(x - x_0)^2 + \rho^2} \quad (3.13)$$

which does approach pure gauge at infinity with winding number one and is non-singular. Here $\Sigma_{\mu\nu}$ is the self-dual analog of $\bar{\Sigma}_{\mu\nu}$.

We also want to consider the generalization of these solutions to super Yang-Mills theory. In 3 + 1 dimensions the super Yang-Mills action in component form is obtained by adding the fermion action

$$\mathcal{L}_F = \text{Tr}(\bar{\chi}^i \gamma^\mu D_\mu \chi) \quad (3.14)$$

to the Minkowski version of (3.1) where χ is a Majorana fermion in the adjoint representation of the gauge group. The Euclidean continuation of this action is not straightforward because of the absence of Majorana fermions in Euclidean space. This can be dealt with in various ways. Luckily these subtleties will not concern us because we will eventually embed the four-dimensional Euclidean theory in a ten-dimensional Minkowski theory which does have Majorana fermions. We can thus proceed as usual except that we must remember that left and right-handed fermions are no longer related by complex conjugation.

To determine whether the instanton solution (3.13) is supersymmetric we need to know whether the supersymmetric variation of this solution is non-zero. The supersymmetry variation of a configuration with vanishing fermion fields is

$$\delta\chi = F_{\mu\nu} \gamma^{\mu\nu} \epsilon \quad (3.15)$$

where ϵ is the parameter of the supersymmetry variation. Under $SO(4)$ ϵ can be decomposed into irreducible representations $\epsilon = \epsilon_L + \epsilon_R$ with $\epsilon_L \sim (2, 1)$ and $\epsilon_R \sim (1, 2)$ under $SO(4) \equiv SU(2)_L \times SU(2)_R$.

For an instanton with $F_{\mu\nu} = \tilde{F}_{\mu\nu}$ we have

$$\delta\chi = F_{\mu\nu} \gamma^{\mu\nu} (1 + \gamma^5) \epsilon \quad (3.16)$$

which is vanishing for ϵ_L but not for ϵ_R . Just as in the kink example, half of the supersymmetries are broken by the solution while the other half give rise to fermion zero modes. Closer examination shows that of the four fermion zero modes predicted by the Atiyah-Singer index theorem for $k = 1$, only two arise from supersymmetry transformations corresponding to the broken supersymmetries as described above. However, at the classical level, this theory is also conformally invariant and this leads to superconformal transformations of the instanton background which produce the two remaining fermion zero modes.

For general k the index theorem predicts $4k$ zero modes, only 4 of which can be constructed as above. There is however an important pairing of fermion zero modes with the bosonic moduli which holds for arbitrary k . Given a variation $\delta A_\mu \in H^1$ we can construct a fermion zero mode of the form

$$\chi_0 = \delta A_\mu \gamma^\mu \epsilon_L \quad (3.17)$$

There is a redundancy in this pairing so that two bosonic zero modes are associated to every fermion zero mode. For details see [9].

4. Magnetic Monopoles and the Montonen-Olive Conjecture

Our final field theory example before discussing string theory has to do with magnetic monopole solutions in Yang-Mills theory and in particular with the properties of monopoles in $N = 4$ super Yang-Mills theory.

The action for a Yang-Mills Higgs theory with gauge group $SU(2)$ and a Higgs field Φ in the adjoint representation is given by

$$S = \int -\frac{1}{2}\text{Tr}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}\text{Tr}D^\mu\Phi D_\mu\Phi - V(\Phi). \quad (4.1)$$

We assume that V has a minimum in which $\langle\text{Tr}\Phi^2\rangle = v^2$ which breaks $SU(2)$ down to $U(1)$. For a static configuration the energy is given by

$$E = \int d^3x \text{Tr}(B^i \pm D^i\Phi)^2 + \int d^3x V(\Phi) + 4\pi v|Q_M| \quad (4.2)$$

where v is the asymptotic vacuum expectation value of Φ . The energy thus satisfies a Bogomolnyi bound

$$E \geq 4\pi v|Q_M| \quad (4.3)$$

where

$$Q_M = \int \vec{B} \cdot d\vec{S} \quad (4.4)$$

is the magnetic charge and $\vec{B}$ above is the asymptotic value of the gauge invariant $U(1)$ field strength. The magnetic charge Q_M is also related through the equations of motion to the element of $\pi_2(SU(2)/U(1)) = Z$ which labels the topological charge carried by the Higgs field configuration. The Bogomolnyi bound is saturated iff $V(\phi) \equiv 0$ and $B^i = \pm D^i\Phi$. If we work in this limit of vanishing potential then we are free to impose as a boundary condition that the vacuum expectation value of Φ approach an arbitrary constant v at spatial infinity. Of course quantum mechanically we expect a potential to be generated by renormalization even if it is absent classically.

Nonetheless let us work in this limit for the time being. Then the first order equations $B^i = \pm D^i\Phi$ can be integrated to give explicit monopole solutions. The charge one solution can be found in [10], multi-monopole solutions are discussed in [11]. In this limit there is a universal formula for the classical mass of the particles of the theory given by

$$M^2 = (4\pi)^2 v^2 (Q_E^2 + Q_M^2) \quad (4.5)$$

where Q_E and Q_M are the electric and magnetic charges of the particle respectively. For monopoles this is just the statement that the Bogomolnyi bound is saturated. For the massive gauge bosons it is just the usual relation between the gauge boson mass and the Higgs vacuum expectation value. The photon is of course neutral and massless as is the remaining physical Higgs boson due to the vanishing potential.

There are some other curious features of this limit. One is the fact that the static force between two monopoles of like charge or between two gauge bosons of like charge vanishes. This is due to a cancellation between a repulsion due to photon exchange and an attraction due to massless Higgs boson exchange [12].

These facts as well as some others led Montonen and Olive [13], following ideas of [14] to conjecture the existence of a "dual" formulation of the theory in which electric and magnetic charges are exchanged and in which the magnetic monopoles become the

gauge bosons and the gauge bosons arise as solitons of the dual theory. In two spacetime dimensions the Thirring model – sine-Gordon duality provides an example where topological and Noether charges are exchanged in the theory and its dual, but this duality relies on the peculiar properties of two dimensions.

One stumbling block to this more ambitious conjecture in $3 + 1$ dimensions is the fact that monopoles appear to have spin zero while gauge bosons of course have spin one. In addition, the vanishing of the potential is not natural from the quantum point of view so there is no reason to expect the mass formula to be exact when quantum corrections are included.

This second objection can be overcome by embedding the theory in $N = 2$ super Yang-Mills theory. Then the charges Q_E and Q_M appear as central charges in the supersymmetry algebra as in the simpler kink example and the mass formula (4.5) is exact for supersymmetric states [4]. However an inspection of the fermion zero modes and the resulting monopole spectrum in this theory shows that the monopole states fill out a matter supermultiplet consisting of spin zero and spin one-half states. In order to construct monopoles with integer spin it is necessary to extend the supersymmetry to $N = 4$, the maximal allowable global supersymmetry in $3 + 1$ dimensions. This theory has a number of remarkable features. First, the structure of the fermion zero modes is such that the monopole supermultiplet now coincides with the gauge supermultiplet and includes states of spin 1, $1/2$, and 0 [15]. Second, the scalar potential has exact flat directions due to supersymmetry and again the mass formula is exact. Finally, this theory is finite with vanishing beta-function so that a duality which relates $g \rightarrow 1/g$ can make sense quantum mechanically at all scales. Thus in this special theory all of the simple objections to the existence of the sort of duality suggested by Montonen and Olive disappear! Of course this is a far cry from showing that such a duality actually holds, but the evidence is suggestive enough that the idea is well worth pursuing.

Finally, it is perhaps worth mentioning that the $N = 4$ theory and the Montonen-Olive conjecture may have some ties to ten-dimensional physics and hence to string theory. For one thing, the $N = 4$ theory can be obtained by dimensional reduction of $N = 1$ super Yang-Mills in ten dimensions. Second, the Montonen-Olive conjecture for more general gauge groups says that the dual gauge group should have a weight lattice dual to the weight lattice of the original group. Self-dual lattices of course play a crucial role in ten-dimensional heterotic string theory with the gauge groups $SO(32)/Z_2$ and $E_8 \times E_8$ with self-dual lattices being singled out by anomaly cancellation. Finally, a stringy analog of the Montonen-Olive conjecture[2] will be briefly discussed in the following section.

5. Low-Energy Heterotic String Theory

Having discussed soliton and instanton solutions of various supersymmetric field theories we would like to generalize these considerations to string theory. Let us first discuss the problem of finding string solitons via the “strings in background fields” spacetime approach. The beta functions for strings propagating in a background of massless fields are the equations of motion of a certain master spacetime action which can be computed as an expansion in the string tension α' . For the heterotic string, the leading terms in this

action are identical to the $D = 10$, $N = 1$ supergravity and super Yang-Mills action. The bosonic part of this action reads

$$S = \frac{1}{\alpha'^4} \int d^{10}x \sqrt{-g} e^{-2\phi} \left(R + 4(\nabla\phi)^2 - \frac{1}{3}H^2 - \frac{\alpha'}{30} \text{Tr} F^2 \right), \quad (5.1)$$

where the three-form antisymmetric tensor field strength is related to the two-form potential by the familiar anomaly equation [16]

$$H = dB + \alpha' \left(\omega_3^L(\Omega_-) - \frac{1}{30} \omega_3^{YM}(A) \right) + \dots \quad (5.2)$$

(where ω_3 is the Chern-Simons three-form) so that

$$dH = \alpha' (tr R \wedge R - \frac{1}{30} Tr F \wedge F). \quad (5.3)$$

The trace is conventionally normalized so that $Tr F \wedge F = \sum_i F^i \wedge F^i$ with i an adjoint gauge group index. An important, and potentially confusing, point is that the connection $\Omega_\pm$ appearing in (5.2) is a non-Riemannian connection related to the usual spin connection ω by

$$\Omega_{\pm M}^{AB} = \omega_M^{AB} \pm H_M^{AB}. \quad (5.4)$$

Since the antisymmetric tensor field plays a crucial role in all of our solutions, this subtlety will be crucial.

Rather than directly solve the equations of motion for this action, it is much more convenient to look for bosonic backgrounds which are annihilated by *some* of the $N=1$ supersymmetry transformations (only the vacuum is annihilated by all the the supersymmetries). Both the kink solution (2.6) and the self-dual equation (3.4) could have been found in this manner. The Fermi field supersymmetry transformation laws which follow from (5.1) are

$$\begin{aligned} \delta\chi &= F_{MN} \gamma^{MN} \epsilon \\ \delta\lambda &= (\gamma^M \partial_M \phi - \frac{1}{6} H_{MNP} \gamma^{MNP}) \epsilon \\ \delta\psi_M &= (\partial_M + \frac{1}{4} \Omega_{-M}^{AB} \gamma_{AB}) \epsilon, \end{aligned} \quad (5.5)$$

and it is apparent that to find backgrounds for which all of (5.5) vanish, it is only necessary to solve first-order equations, rather than the more complicated second-order equations which follow from varying the action. We will shortly construct a simple ansatz for the bosonic fields which does just this.

Although there are a variety of soliton or instanton solutions of (5.5) we could consider, it is useful to first study the simplest and most natural solutions. In realistic theories we would consider a spacetime of the form $M \times K$ with M four-dimensional Minkowski or Euclidean space and K a six-dimensional compact space which solves the equations of motion following from (5.1). Soliton and instanton solutions involving gauge fields on

such spaces have been studied and a general classification has been discussed in [17]. However these solutions are either related to the usual Yang-Mills instantons or involve the structure of K in various ways. A simpler but less realistic starting point would be to ask for solutions to string theory in ten-dimensional Minkowski space (or Euclidean space in the case of instantons). We will discuss solutions of this form both because they allow us to discuss solitons without the complication of discussing compactification and because the solitons we find may have implications for fundamental string theory.

The solutions of (5.5) that we will consider are fivebrane solitons. The motivation for making such an apparently arbitrary choice is based on an analogy between these solitons in string theory and Dirac monopoles in electromagnetism. In $3+1$ dimensions a point particle naturally couples to a one-form potential A via

$$\int_L A \equiv \int d\tau \frac{dx^\mu}{d\tau} A_\mu \quad (5.6)$$

where L indicates the world-line of the particle, determined by $x_\mu(\tau)$ which gives the particle location in spacetime as a function of a time-like parameter τ . This coupling results in an “electric” source term in the equation of motion for the field strength $F = dA$

$$d^*F = j_E. \quad (5.7)$$

Using Stoke’s theorem the electric charge of the point particle is

$$Q_E = \int_{S_\infty^2} {}^*F \quad (5.8)$$

where S_∞^2 is a two-sphere at spatial infinity.

Dirac conjectured that there might exist magnetic monopoles which would obey a dual version of these equations

$$dF = 0 \implies dF = \tilde{j}_M \quad (5.9)$$

$$d^*F = j_E \implies d^*F = 0 \quad (5.10)$$

$$Q_E = \int_{S_\infty^2} {}^*F \implies Q_M = \int_{S_\infty^2} F \quad (5.11)$$

and showed that quantum consistency imposes a quantization condition on the product of the electric and magnetic charges $Q_E Q_M = 2\pi n$ with n an integer.

There is a construction analogous to this in string theory. Since we now want to couple to the world-sheet of a string rather than the world-line of a particle we introduce a two-form potential B , called a Kalb-Ramond field, which couples to the string world-sheet S via $\int_S B$ and has a three-form field strength $H = dB$. This is the same H as in (5.2) except that for the moment we ignore the Chern-Simons terms in Equation (5.2). The analogs of the previous equations and their dual forms are then (in D spacetime dimensions)

$$dH = 0 \implies dH = j_M \quad (5.12)$$

$$d^*H = j_E \implies d^*H = 0 \quad (5.13)$$

$$Q_E = \int_{S_{\infty}^{D-3}} *H \implies Q_M = \int_{S_{\infty}^3} H. \quad (5.14)$$

Because $*H$ is a $D-3$ form, it must be integrated over a $D-3$ dimensional manifold. This fits perfectly with our string picture since a static infinite string in D spacetime dimensions has a $D-3$ sphere at spatial infinity. However this also means that the dual object has a charge measured by integrating H over a three-dimensional manifold. So, unless $D = 6$, the string and its dual object do not have the same dimension. For example, consider $D = 4$. It is well known that the two form B is equivalent to a massless scalar field in four dimensions and that this scalar field has couplings analogous to the usual axion field. The “electric” charge, measured by integrating $*H$ over the S^1 at infinity around an infinite string, is just the axion charge of the string. The dual object has a “magnetic” charge obtained by integrating H over a three-sphere at infinity. But this means that the dual object must be localized in both space and time, i.e. it is an instanton! Thus we learn the important fact that in four dimensions strings are dual, in the Dirac sense, to instantons. If we move up to $D = 10$ as required by superstring theory, then the dual object is a fivebrane, that is an object with extent in five spatial dimensions which sweeps out a six-dimensional world-volume as it evolves in time.

This fivebrane is also intimately related to Yang-Mills instantons in four dimensions. One way to see this connection is to note that in string theory we no longer have $H = dB$ but rather

$$H = dB - \frac{\alpha'}{30} \omega_3^{YM} \quad (5.15)$$

(the additional gravitational contributions which will be discussed shortly). We thus have the equations

$$dH = -\frac{\alpha'}{30} \text{Tr}(F \wedge F) \quad (5.16)$$

$$d^*H = 0 \quad (5.17)$$

in a region free of electric sources. Comparison with the previous equations clearly shows that the Yang-Mills topological charge density acts as a magnetic source term for H .

This duality argument doesn't guarantee that the dual objects actually exist, of course, but we shall see that in fact they do. The existence of objects which are *geometrically* dual to strings is quite interesting and raises the question of whether they could also be *dynamically* dual as in the conjecture of Montonen and Olive. This possibility was raised in [2], some issues associated with and evidence for this conjecture are discussed in [18]. Although the idea is a tantalizing one, proving or disproving it seems beyond our current ability.

With this motivation let us attempt to construct a fivebrane solution to (5.5). The supersymmetry variations are determined by a positive chirality Majorana-Weyl $SO(9,1)$ spinor ϵ . Because of the fivebrane structure, it is useful to note that ϵ decomposes under $SO(9,1) \supset SO(5,1) \otimes SO(4)$ as

$$16 \rightarrow (4_+, 2_+) \oplus (4_-, 2_-) \quad (5.18)$$

where the $\pm$ subscripts denote the chirality of the representations. Denote world indices of the four-dimensional space transverse to the fivebrane by $\mu, \nu = 6 \dots 9$ and the corresponding tangent space indices by $m, n = 6 \dots 9$. We assume that no fields depend on the longitudinal coordinates (those with indices $M = 0 \dots 5$) and that the nontrivial tensor fields in the solution have only transverse indices. Then the gamma matrix terms in (5.5) are sensitive only to the $SO(4)$ part of ϵ and, in particular, to its $SO(4)$ chirality.

One immediately sees how to make the gaugino variation vanish (in what follows we treat ϵ as an $SO(4)$ spinor and let all indices be four-dimensional): As a consequence of the four-dimensional gamma-matrix identity $\gamma^{mn}\epsilon_{\pm} = \mp \frac{1}{2}\epsilon^{mnr s}\gamma^{rs}\epsilon_{\pm}$ one has $F_{mn}\gamma^{mn}\epsilon_{\pm} = \mp \tilde{F}_{mn}\gamma^{mn}\epsilon_{\pm}$, where the dual field strength is defined by $\tilde{F}_{mn} = \frac{1}{2}\epsilon_{mnr s}F^{rs}$. Therefore, $\delta\chi$ vanishes if $F_{mn} = \pm\tilde{F}_{mn}$ and $\epsilon = (4_{\pm}, 2_{\pm})$ which is to say that if the gauge field is taken to be an *instanton*, then $\delta\chi$ vanishes for all supersymmetries with positive $SO(4)$ chirality. This is just the argument of (3.16) except that the supersymmetry parameter now transforms under $SO(5, 1)$ as well.

To deal with the other supersymmetry variations, we must adopt an *ansatz* [22] for the non-trivial behavior of the metric and antisymmetric tensor fields in the four dimensions transverse to the fivebrane. For the metric tensor we write

$$g_{\mu\nu} = e^{+2\phi}\delta_{\mu\nu}, \quad \mu, \nu = 6 \dots 9 \quad (5.19)$$

and for the antisymmetric tensor field strength

$$H_{\mu\nu\lambda} = -\epsilon_{\mu\nu\lambda}{}^{\sigma}\partial_{\sigma}\phi, \quad (5.20)$$

where ϕ is to be identified with the dilaton field. With this ansatz and the rather obvious vierbein choice $e_{\mu}^m = \delta_{\mu}^m e^{+\phi}$, we can also calculate the generalized spin connection (5.4) which appears in (5.5) and (5.2):

$$\Omega_{\pm\mu mn} = \delta_{m\mu}\partial_n\phi - \delta_{n\mu}\partial_m\phi \mp \epsilon_{\mu mn}{}^{\rho}\partial_{\rho}\phi. \quad (5.21)$$

Now consider the $\delta\lambda$ term in (5.5). Because of the ansatz, both terms are linear in $\partial\phi$. By standard four-dimensional gamma-matrix algebra, the relative sign of the two terms is proportional to the $SO(4)$ chirality of the spinor ϵ . We have chosen the sign and normalization of the ansatz for H so that $\delta\lambda$ vanishes for $\epsilon \in (4_{+}, 2_{+})$. Finally, consider the gravitino variation in (5.5). A crucial fact, following from (5.21), is that while $\Omega_{\pm}$ would in general be an $SO(4)$ connection, with the chosen ansatz it is actually pure $SU(2)$. To be precise,

$$\Omega_{\pm\mu}{}^{mn}\gamma_{mn}\epsilon_{\eta} = 2(\gamma_{\mu}{}^{\rho}\partial_{\rho}\phi)(1 \mp \eta 1)\epsilon_{\eta}, \quad (5.22)$$

(where $\eta = \pm$) so that $\Omega_{\pm}$ annihilates the $(4_{\pm}, 2_{\pm})$ spinor. Since (5.5) involves only Ω_{-} , it suffices to take ϵ to be a *constant* $(4_{+}, 2_{+})$ spinor to make the gravitino variation vanish. It can be shown that *all* supersymmetric fivebrane configurations are of the form given by the ansatz (5.19), (5.20).

Putting all this together, we see that if we choose the gauge field to be any instanton and fix the metric and antisymmetric tensor in terms of the dilaton according to the

above ansatz, then the state is annihilated by all supersymmetry variations generated by a spacetime constant $(4_+, 2_+)$ spinor. Thus, *half* of the supersymmetries are unbroken, and the other half, by standard reasoning, will be associated with fermionic zero-modes bound to the soliton.

The one unresolved question concerns the functional form of the dilaton field. Notice that the ansatz for the antisymmetric tensor was given in terms of its three-form field strength $H_{\mu\nu\rho}$, rather than its two-form potential $B_{\mu\nu}$. This is potentially inconsistent, since the curl of the field strength must satisfy the anomalous Bianchi identity (5.3). Within the ansatz (5.20), the curl of H is given by

$$dH = -\frac{1}{2} \star \square e^{-2\phi} = \alpha' (\text{tr} R \wedge R - \frac{1}{30} \text{Tr} F \wedge F) . \quad (5.23)$$

where $\square$ is the flat space laplacian and $\star$ is the four-dimensional Hodge dual. This equation can be solved perturbatively in α' beginning with the instanton solution for F . Equation (5.23) then implies that $\partial\phi \sim O(\alpha')$ which in turn implies that $R \sim O(\alpha')$. Therefore, to leading order in α' , one is entitled to drop the $R \wedge R$ term in (5.23). Substituting the explicit gauge field strength for an instanton of scale size ρ , one obtains the following dilaton solution:

$$e^{+2\phi} = e^{+2\phi_0} + 8\alpha' \frac{(x^2 + 2\rho^2)}{(x^2 + \rho^2)^2} + O(\alpha'^2) . \quad (5.24)$$

The metric and antisymmetric tensor fields are built out of this dilaton field according to the spacetime-supersymmetric ansatz of (5.19) and (5.20). One can examine higher-order in α' corrections to the beta functions and verify that the solution must receive corrections. At the same time, one can examine subleading corrections to the supersymmetry transformations [23] and verify that it is possible to maintain spacetime supersymmetry in the α' -corrected solution. However one would like to know if there is an *exact* solution of string theory which agrees with this one at long distances. In fact the existence of such an exact solution can be demonstrated using low-energy spacetime supersymmetry. As these notes focus on the worldsheet point of view, we refer the reader to the original literature [18] for details.

While (5.24) is in some sense the most obvious supersymmetric fivebrane solution, there is in fact a different solution with more worldsheet supersymmetry. This extra worldsheet supersymmetry will be seen to have powerful consequences: the solution is exact without any α' corrections. This symmetric solution is characterized, from the spacetime point of view, by $dH = 0$. This condition requires, according to (5.3), that the curvature $R(\Omega_-)$ should cancel against the instanton Yang-Mills field F . We will take the instanton to be embedded in an $SU(2)$ subgroup of the gauge group (this is always the lowest-action instanton), so what is needed is that Ω_- be a self-dual $SU(2)$ connection. In fact Ω_- is manifestly an $SU(2)$ connection, so the only issue is self-duality. Given the special ansatz and coordinate system of (5.19), it is easy to calculate the curvature of $\Omega_\pm$. It is then a simple arithmetical exercise to show that, in four dimensions and under the condition that $\square e^{2\phi} = 0$, Ω is self-dual:

$$R(\Omega_\pm)_{\mu\nu}^{mn} = \mp \frac{1}{2} \epsilon_{\mu\nu}{}^{\lambda\sigma} R(\Omega_\pm)_{\lambda\sigma}^{mn} . \quad (5.25)$$

Since a self-dual $SU(2)$ connection is an instanton connection, it will be possible to choose a gauge instanton which exactly matches the “metric” instanton Ω_- and makes the *r.h.s* of (5.3) vanish, thus making the whole solution self-consistent. In the next section, we will explore the qualitative properties of the solutions which we have constructed in the above rather roundabout manner.

A feature of the above development which could cause confusion is the intricate interplay of the two non-Riemannian connections $\Omega_{(\pm)}$. To refresh the reader’s memory, we will summarize the essentials of this phenomenon (we denote the $(4_+, 2_+)$ spinor by ϵ_+): The gravitino supersymmetry variation equation boils down to $\Omega_-^{ab} \gamma_{ab} \epsilon_+ = 0$ which in turn implies that $R_{\mu\nu}(\Omega_-)^{ab} \gamma_{ab} \epsilon_+ = 0$. The index-pair interchange symmetry $R(\Omega_+)_{ab,cd} = R(\Omega_-)_{cd,ab}$, allows us to convert the previous condition for ϵ_+ to $R_{\mu\nu}(\Omega_+)^{ab} \gamma^{\mu\nu} \epsilon_+ = 0$. If we then make the identification $F_{\mu\nu}{}^{ab} \sim R_{\mu\nu}(\Omega_+)^{ab}$, we see that we have reproduced the gaugino supersymmetry variation equation. This is simply to emphasize that, because of the crucial role of the antisymmetric tensor in these solutions, the precise way in which the Ω_+ ’s and Ω_- ’s appear in the various equations we deal with is tightly constrained and quite critical.

6. Development and Interpretation of the Solutions

Now we will work out the geometry and physical interpretation of the solutions described in the previous section. We begin with a discussion of the “gauge” solution described by (5.24).

There are two charges one can associated with this solution. These are the instanton winding number

$$\nu = \frac{1}{480\pi^2} \int \text{Tr} F \wedge F, \quad (6.1)$$

where the integral is over a four-dimensional cross section, and the axion charge

$$Q = -\frac{1}{2\pi^2} \int H, \quad (6.2)$$

where the integral is over an asymptotic S^3 surrounding the fivebrane. The instanton winding number is quantized as usual and the minimal value is $\nu = 1$. The axion charge Q is also quantized in integer multiples of α' . The point is that, if the flux of H through a non-trivial S^3 is non-zero, then there cannot be a unique two-form potential B covering the whole sphere. The best one can do is to have two sections, $B^\pm$, covering the upper and lower halves of the S_3 and related to each other by a gauge transformation in an overlap region which is topologically an S_2 . Since the sigma model action involves B , not H , the non-uniqueness of B could lead to an ill-defined sigma model path integral. It is possible to show that, with our definitions of the sigma model action, this danger is avoided if and only if the flux of H is an integral multiple of α' . The details of this argument can be found in [24]. Although the minimal allowed value of Q is α' , the gauge solution in fact has $Q = 8\alpha'$ as can be verified by explicit calculation.

The form of the solution (5.24) was determined by a “low-energy” expansion with the expansion parameter being α'/ρ^2 . Solutions of scale size less than $\sqrt{\alpha'}$ can be shown to exist using arguments based on low-energy supersymmetry, but are not well approximated by the solution of (5.24).

We now turn to the “symmetric” solution. In the previous section we found that the following is a solution of the low-energy spacetime effective action of the heterotic string:

$$\begin{aligned} ds^2 &= e^{2\phi(x)} \delta_{\mu\nu} dx^\mu dx^\nu + \eta_{\alpha\beta} dy^\alpha dy^\beta \\ H_{\mu\nu\lambda} &= -\epsilon_{\mu\nu\lambda}{}^\sigma \partial_\sigma \phi \\ F_{\mu\nu}{}^{mn} &= \tilde{F}_{\mu\nu}{}^{mn} = R_{\mu\nu}(\Omega_-)^{mn}, \end{aligned} \quad (6.3)$$

where $\mu, \nu = 6 \dots 9$, $\alpha, \beta = 0 \dots 5$ and $\eta_{\alpha\beta}$ is the Minkowski metric. The last equation expresses the fact that the gauge field is a self-dual instanton with moduli chosen so that it coincides (up to gauge transformations of course) with the curvature of the generalized connection of the theory. The consistency condition for all this is just $\square e^{2\phi} = 0$.

The solution of the consistency condition on ϕ is just a constant plus a sum of poles:

$$e^{2\phi} = e^{2\phi_0} + \sum_{i=1}^N \frac{Q_i}{(x - x_i)^2} \quad (6.4)$$

which should be compared with the analogous expression (3.11) which appeared in the 'tHooft ansatz. The constant term is fixed by the (arbitrary) asymptotic value of the dilaton field, ϕ_0 . In string theory, e^ϕ is identified with the local value of the string loop coupling constant, g_{str} . For the solution described by (6.4), g_{str} goes to a constant at spatial infinity and goes to infinity at the locations of the poles! We shall worry about the physical interpretation of this fact in due course. Now, the metric of our solution is conformally flat with conformal factor given by (6.4). Since ϕ goes to a constant at infinity, the geometry is asymptotically flat, which is precisely what we want for a soliton interpretation. In the neighborhood of a singularity, we can replace $e^{2\phi}$ by a simple pole Q/r^2 and obtain the approximate line element

$$\begin{aligned} ds^2 &\sim \frac{Q}{r^2} (dr^2 + r^2 d\Omega_3^2) \\ &= dt^2 + Q d\Omega_3^2 \end{aligned} \quad (6.5)$$

where $d\Omega_3^2$ is the line element on the unit three-sphere and we have introduced a new radial coordinate $t = \sqrt{Q} \log(r/\sqrt{Q})$. This expression becomes more and more accurate as $t \rightarrow -\infty$. In this same limit, the other fields are given by

$$\phi = -t/\sqrt{Q} \quad H = -Q\epsilon_3, \quad (6.6)$$

where ϵ_3 is the volume form on the three sphere. In Sect. 5 we will see that the linear behavior of the dilaton field plays a crucial role in the underlying exact conformal field

theory. The geometry described by (6.5) is a cylinder whose cross-section is a three-sphere of constant area $2\pi^2 Q$. The global geometry is that of a collection of semi-infinite cylinders, or semi-wormholes (one for each pole in $e^{+2\phi}$), glued into asymptotically flat four-dimensional space. The semi-wormholes are semi-infinite since the approximation of (6.5) becomes better and better as $t \rightarrow -\infty$ and breaks down as $t \rightarrow +\infty$. It is these semi-wormholes which we propose to interpret as solitons.

A further crucial fact is that the residues, Q , are quantized. Consider an S_3 which surrounds a single pole, of residue Q , in $e^{+2\phi}$. The net flux of H through this S_3 is entirely due to the enclosed pole and can easily be calculated:

$$\int_{S_3} H = -2\pi^2 Q \quad (6.7)$$

The consequence for us is that the residues of the poles in $e^{2\phi}$ are discretely quantized: $Q_i = n_i \alpha'$. As a result, the cross-sectional areas of the individual semi-wormholes are quantized in units of $2\pi^2 \alpha'$ and there is thus a minimal transverse scale size of the fivebrane. (This fact may be useful in future attempts to quantize the transverse fluctuations of the fivebrane.)

Finally, we want to characterize the instanton component of this solution. The key point is that, when the dilaton field satisfies $\square e^{2\phi} = 0$, we can construct a self-dual $SU(2)$ connection out of the scalar field ϕ and we want to identify this connection with the gauge instanton.

But this is simple in terms of the ansatz (3.8) and the corresponding scalar field (3.11). An important point is that the total instanton number of the solution built on f of (3.11) is N , the number of poles. The gauge potential which follows from taking f to have a single pole is

$$A_\mu = -2\rho^2 \tilde{\Sigma}_{\mu\nu} \frac{x^\nu}{x^2(x^2 + \rho^2)}, \quad (6.8)$$

an expression which one immediately recognizes as the singular gauge instanton of scale size ρ centered at $x = 0$. The only way this can match our construction of a self-dual generalized connection is if we make the identification

$$f(x) = e^{-2\phi_0} e^{2\phi} = 1 + \sum_{i=1}^N \frac{e^{-2\phi_0} Q_i}{(x - x_i)^2}. \quad (6.9)$$

Thus, given the solution (6.4) for the dilaton field, we can assert that the associated instanton has instanton number N , with instantons of scale size $\rho_i^2 = e^{-2\phi_0} Q_i \alpha'$ localized at positions x_i . Since the Q_i are quantized, so are the instanton scale sizes. The only free parameters (moduli) are the $4N$ center locations of the instantons. In ten dimensions, the multiple instanton solution corresponds to multiple fivebranes with the locations in the transverse four-dimensional space of the individual fivebranes given by the center coordinates of the individual instantons.

An important fact about the solution we have just constructed is that it is not perturbative in α' . As we saw in the discussion following (6.3), the semi-wormhole associated

with a pole of residue $Q = n\alpha'$ has a cross-section which is a sphere of area $2\pi^2 Q$ and therefore has curvature $R \sim 1/Q$. In the perturbative sigma model approach to strings in background fields, one finds that the sigma model expansion parameter is $\alpha'R$. In the case at hand, this becomes $\alpha'/Q = 1/n$, which is obviously not small for the elementary fivebrane, which has $n = 1$. Since our solution has been constructed by solving the leading-order-in- α' beta function equations, ignoring all higher-order corrections, one can legitimately worry whether it makes any sense. In the next two sections we will present evidence that all corrections to this particular solution actually vanish, and the leading-order solution is exact.

There is another perturbation theory issue which should be mentioned here. String theory has two expansion parameters: the string tension α' and the string loop coupling constant $g_{str} \sim e^{\phi_0}$. The latter is the quantum expansion parameter of string theory and, in this paper, we are working to zeroth-order in an expansion in g_{str} . In effect, we are producing an exact solution of *classical* string field theory. However, as we have already pointed out, our solution has the unusual feature that g_{str} grows without limit down the throat of a semi-wormhole so that there is, strictly speaking, no reliable classical limit! Since virtually nothing is known about non-perturbative-in- g_{str} physics, we don't know what this means for the ultimate validity of this sort of solution. Similar issues arise in the matrix model/Liouville theory approach to two-dimensional quantum gravity, and we hope eventually to gain some insight from that source.

Another interesting point concerns what happens when we lift the requirement of spacetime supersymmetry and look for solutions of the beta function equations rather than the condition that some supersymmetry charges annihilate the solution. Our solutions have the property that the mass (the ADM mass, to be precise) per unit fivebrane area is proportional to the axion charge: $M_5 = Q$. This equality can be understood via a Bogomolnyi bound: any solution of the leading-order field equations with the fivebrane topology must satisfy the inequality $M_5 \geq Q$ and our solution saturates the inequality. One can easily imagine a process in which mass, but not axion charge, is increased by sending a dilaton wave down one of the semi-wormhole throats. Since the semi-wormhole throat is semi-infinite, this wave need not be reflected back: It can continue to propagate down the throat forever, leaving an exterior solution for which $M_5 > Q$. Such solutions of the leading-order beta function equations have indeed been found [25] and they resemble the familiar Reissner-Nordstrom family of charged black holes: they have an event horizon and a singularity, but the singularity retreats to infinity as the mass is decreased to the extremal value that saturates the Bogomolnyi bound. Perhaps not surprisingly, the non-extremal solutions are not annihilated by any spacetime supersymmetries. Nevertheless it is possible in some cases to give exact conformal field theory constructions of these solutions [26]. These developments should eventually allow us to make progress on understanding the string physics of black holes, Hawking radiation and the like.

In the rest of these lectures, we will pursue the much more limited goal of showing that the symmetric solution is an exact solution of string theory using world-sheet arguments.

7. Worldsheet Sigma Model Approach

To show conclusively that a given spacetime configuration is a solution of string theory,

we must show that it derives from an appropriate superconformal worldsheet sigma model. In this section we will show that the worldsheet sigma models corresponding to the fivebranes constructed in section 5 possess extended worldsheet supersymmetry of type (4,4). The notation derives from the fact that in a conformal field theory, the left-moving fields (functions of z) and the right-moving fields (functions of $\bar{z}$) are dynamically independent. It is therefore possible to have different numbers of right- and left-moving supercharges $Q_{\pm}^I$. The general case, referred to as (p, q) supersymmetry, is described by the algebra

$$\begin{aligned} \{Q_+^I, Q_+^J\} &= 2\delta^{IJ} P_+ & I, J &= 1 \dots p \\ \{Q_-^I, Q_-^J\} &= 2\delta^{IJ} P_- & I, J &= 1 \dots q \\ \{Q_+^I, Q_-^J\} &= 0 \end{aligned} \quad (7.1)$$

The minimal possibility, corresponding to a generic solution of the heterotic string, has $(1, 0)$ supersymmetry. Any left-right-symmetric, and therefore non-anomalous theory, will have (p, p) supersymmetry (this is sometimes referred to as $N = p$ supersymmetry). The maximal possibility is $(4, 4)$ which, it turns out, is what is realized in our fivebrane solution. We will argue that, in the $(4, 4)$ case, there is a nonrenormalization theorem which makes the lowest-order in α' solution for the spacetime fields exact. The latter issue is closely related to the question of finiteness of sigma models with torsion and with extended supersymmetry [27, 28] and the results we find are slightly at variance with the conventional wisdom, at least as we understand it. We will comment upon this at the appropriate point.

First we digress to explain why we expect four-fold extended supersymmetry in this problem. The models of interest to us are structurally equivalent to a compactification of ten-dimensional spacetime down to six dimensions: there are six flat dimensions (along the fivebrane) described by a free field theory and four 'compactified' dimensions (transverse to the fivebrane) described by a nontrivial field theory. The fact that the 'compactified' space is not really compact has no bearing on the supersymmetry issue. The defining property of all the fivebranes of section 5 is that they are annihilated by the generators of a six-dimensional $N=1$ spacetime supersymmetry. That is, they provide a compactification to six dimensions which maintains $N=1$ spacetime supersymmetry. Now, it is well-known that in compactifications to four dimensions, the sigma model describing the six compactified dimensions must possess $(2, 0)$ worldsheet supersymmetry in order for the theory to possess $N=1$ four-dimensional spacetime supersymmetry [29]. Roughly speaking, the conserved $U(1)$ current of the $(2, 0)$ superconformal algebra defines a free boson which is used to construct the spacetime supersymmetry charges. It is also known that, if one wants to impose $N=2$ four-dimensional spacetime supersymmetry, the compactification sigma model must have $(4, 0)$ supersymmetry [30]. The conserved $SU(2)$ currents of the $(4, 0)$ superconformal algebra are precisely what are needed to construct the larger set of $N=2$ spacetime supersymmetry charges. Since, by dimensional reduction, $N=1$ supersymmetry in six dimensions is equivalent to $N=2$ in four dimensions, the above line of argument implies that spacetime supersymmetric compactifications to six dimensions (including our fivebrane) require a compactification sigma model with at least $(4, 0)$ worldsheet supersymmetry. Since our solution is constructed to cancel the anomaly, it will be left-right symmetric and therefore automatically of type $(4, 4)$.

Now we turn to a study of string sigma models. The generic sigma model underlying the heterotic string describes the dynamics of D worldsheet bosons X^M and D right-moving worldsheet fermions ψ_R^M (where D, typically ten, is the dimension of spacetime) plus left-moving worldsheet fermions λ_L^a which lie in a representation of the gauge group G (typically $SO(32)$ or $E_8 \otimes E_8$). The generic Lagrangian for this sigma model is written in terms of coupling functions G_{MN} , B_{MN} and A_M which eventually get interpreted as spacetime metric, antisymmetric tensor and Yang-Mills gauge fields. This Lagrangian has the explicit form [31]

$$\frac{1}{4\pi\alpha'} \int d^2\sigma \{ G_{MN}(X) \partial_+ X^M \partial_- X^N + 2B_{MN}(X) \partial_+ X^M \partial_- X^N \\ + iG_{MN} \psi_R^M \mathcal{D}_- \psi_R^N + i\delta_{ab} \lambda_L^a \mathcal{D}_+ \lambda_L^b + \frac{1}{2} (F_{MN})_{ab} \psi_R^M \psi_R^N \lambda_L^a \lambda_L^b \} \quad (7.2)$$

where $H = dB$. In this expression, the covariant derivatives on the left-moving fermions are defined in terms of the Yang-Mills connection, while the covariant derivatives on the right-moving fermions are defined in terms of a non-riemannian connection involving the torsion (which already appeared in section 5):

$$\mathcal{D}_- \psi_R^A = \partial_- \psi_R^A + \Omega_{-N}^A{}_B \partial_- X^N \psi_R^B, \\ \mathcal{D}_+ \lambda_L^a = \partial_+ \lambda_L^a + A_N{}^a{}_b \partial_+ X^N \lambda_L^b. \quad (7.3)$$

We use indices of type M for coordinate space indices, type A for the tangent space and type a for the gauge group. An absolutely crucial feature of this action is that the connection appearing in the covariant derivative of the right-moving fermions is the generalized connection Ω_- , *not* the Christoffel connection. This action has a naive $(1,0)$ worldsheet supersymmetry and can be written in terms of $(1,0)$ superfields. Superconformal invariance is broken by anomalies of various kinds unless the coupling functions satisfy certain ‘beta function’ conditions [32] which are equivalent to the spacetime field equations discussed in section 5. The dilaton enters these equations in a rather roundabout, but by now well-understood, way [31].

To proceed further, we must construct the specific sigma models corresponding to the fivebrane solutions. For the generic fivebrane, (7.2) undergoes a split into a nontrivial four-dimensional theory and a free six-dimensional theory: the sigma model metric (as opposed to the canonical general relativity metric) then describes a flat six-dimensional spacetime times four curved dimensions. The right-moving fermions couple via the kinetic term to the generalized connection Ω_- , which acts only on the four right-movers lying in the tangent space orthogonal to the fivebrane. The other six right-movers are free (we momentarily ignore the four-fermi coupling) so there is a six-four split of the right-movers as well. The left-moving fermions couple to an instanton gauge field which may or may not be identified with the *other* generalized connection, Ω_+ . In all the cases of interest to us, the gauge connection is an instanton connection and acts only in some $SU(2)$ subgroup of the full gauge group, so that four of the left-movers couple nontrivially, while the other 28 are free. Finally, the four-fermion interaction term couples together precisely those left- and right-movers which couple to the nontrivial gauge and Ω_- connections and is therefore

consistent with the six-four split defined by the kinetic terms. The remaining variables can be regarded as defining a heterotic, but free, theory (6 X , 6 ψ_R and 28 λ_L) living in the six ‘uncompactified’ dimensions along the fivebrane. From now on, we focus our attention on the nontrivial piece of (7.2) referring to the four-dimensional part of the split. For string theory consistency, it must have a central charge of 6, which would be trivially true if the connections were all flat, but is far from obvious for a fivebrane.

Now let us further specialize to the sigma model underlying the left-right symmetric (and therefore non-anomalous) fivebrane solution of section 5. It is constructed by identifying the gauge connection with the ‘other’ generalized connection Ω_+ and making that connection self-dual by imposing the condition $\square e^{2\phi} = 0$ on the metric conformal factor. The result of this is that the four bosonic coordinates transverse to the fivebrane and the four nontrivially-coupled left- and right-moving fermions are governed by the worldsheet action

$$\begin{aligned} \frac{1}{4\pi\alpha'} \int d^2\sigma \{ & G_{\mu\nu}(X) \partial_+ X^\mu \partial_- X^\nu + 2B_{\mu\nu}(X) \partial_+ X^\mu \partial_- X^\nu \\ & + iG_{\mu\nu} \psi_R^\mu \mathcal{D}_- \psi_R^\nu + iG_{\mu\nu} \lambda_L^\mu \mathcal{D}_+ \lambda_L^\nu + \frac{1}{2} R(\Omega_+)_{\mu\nu\lambda\rho} \psi_R^\mu \psi_R^\nu \lambda_L^\lambda \lambda_L^\rho \} \end{aligned} \quad (7.4)$$

where $\mathcal{D}_\pm$ are the covariant derivatives built out of the generalized connections $\Omega_\pm$. In fact, as long as the H appearing in $\Omega_\pm$ is given by dB , (7.4) is identical to the basic left-right symmetric, (1,1) supersymmetric nonlinear sigma model with torsion [31]. Despite the apparent asymmetry of the coupling of λ_L to Ω_+ and ψ_R to Ω_- , the theory nonetheless has an overall left-right symmetry (under which $B \rightarrow -B$) and is non-anomalous. To exchange the roles of ψ_R and λ_L one has to replace the curvature of Ω_- by that of Ω_+ . This exchange symmetry property relies on the non-riemannian relation

$$R(\Omega_+)_{\mu\nu\lambda\rho} = R(\Omega_-)_{\lambda\rho\mu\nu} \quad (7.5)$$

which indeed holds for the generalized connection (5.4) when $dH = 0$. To summarize, we have shown that the heterotic sigma model describing the nontrivial four-dimensional geometry of the fivebrane is actually an example of a left-right symmetric sigma model with at least (1,1) supersymmetry. As we will now show, it actually has (4,4) worldsheet supersymmetry.

We now turn to the question of extended supersymmetry. The basic worldsheet supersymmetry of a (1,1) model like (7.4) is

$$\begin{aligned} \delta X^M &= \epsilon_L \psi_R^M + \epsilon_R \psi_L^M \\ \delta \psi_L^A + \Omega_{+M}^A \delta X^M \psi_L^B &= \partial X^A \epsilon_R + \dots \\ \delta \psi_R^A + \Omega_{-M}^A \delta X^M \psi_R^B &= \partial X^A \epsilon_L + \dots \end{aligned} \quad (7.6)$$

The worldsheet supersymmetry of the (1,0) model is obtained by dropping the contributions of ϵ_R and ψ_L . The general structure of a possible second supersymmetry transformation is

$$\begin{aligned} \hat{\delta} X^M &= \epsilon_L f_R(X)_N^M \psi_R^N + \epsilon_R f_L(X)_N^M \psi_L^N \\ \hat{\delta} \psi_L^A + \Omega_{+M}^A \delta X^M \psi_L^B &= -f_L(X)_B^A \partial X^B \epsilon_R + \dots \\ \hat{\delta} \psi_R^A + \Omega_{-M}^A \delta X^M \psi_R^B &= -f_R(X)_B^A \partial X^B \epsilon_L + \dots \end{aligned} \quad (7.7)$$

The function f is normalized and fully defined by the requirements that $\{\hat{\delta}, \delta\} = 0$ and that $\hat{\delta}$ anticommute with itself to give ordinary translations as in (7.1). The question is, what conditions must f satisfy in order for $\hat{\delta}$ to be a symmetry and how many can exist?

This question was first addressed in [33] for the case of left-right symmetric theories without torsion (*i.e.* without an antisymmetric tensor coupling term). The more complex case of left-right symmetry with torsion was subsequently dealt with in [34, 27, 28]. The basic result is that the pair of tensors $f_{R,L}$ must be complex structures, covariantly constant with respect to the appropriate connection:

$$f_{\pm}^2 = -1$$

$$\mathcal{D}_A^{\pm} f_{\pm}^B{}_C = \partial_A f_{\pm}^B{}_C + \Omega_{AD}^{(\pm)B} f_{\pm}^D{}_C - \Omega_{AC}^{(\pm)D} f_{\pm}^B{}_D = 0, \quad (7.8)$$

where the $\pm$ notation is equivalent to the L, R notation. The tensors in (7.8) are written in tangent space indices which is why the generalized spin connections $\Omega^{(\pm)}$ appear in the covariant derivative. The equation could, of course, also have been written in coordinate indices. In general, it is not obvious that such a pair of complex structures can be found, but, if one can, we know that the sigma model actually possesses $(2, 2)$ worldsheet supersymmetry. A further question is whether multiple pairs $f_{\pm}^{(r)}$ of such complex structures can be found. If we can find $p - 1$ of them, then the sigma model has (p, p) supersymmetry. It turns out that the only consistent possibility for multiple complex structures is that there be three of them [33] and that they satisfy the Clifford algebra

$$f_{\pm}^{(r)} f_{\pm}^{(s)} = -\delta_{rs} + \epsilon_{rst} f_{\pm}^{(t)}. \quad (7.9)$$

This corresponds to the case of $(4, 4)$ supersymmetry. It is worth noting that each complex structure leads to a conserved (chiral) current:

$$J_{\pm}^{(r)} = \psi_{\pm}^A (f_{\pm}^{(r)})_{AB} \psi_{\pm}^B. \quad (7.10)$$

This yields a $U(1)$ symmetry in the $(2, 2)$ case and an $SU(2)$ symmetry in the $(4, 4)$ case.

The question of left-right asymmetric theories, such as those which underlie the “gauge” fivebranes discussed in section 5, is more delicate. According to [27], a heterotic sigma model will have $(p, 0)$ supersymmetry if there are $p - 1$ complex structures $f_{+}^{(r)}$ which are covariantly constant under the connection which couples to the right-moving fermions (those which do not couple to the gauge field) and if the gauge field (which affects the left-moving fermions) satisfies a condition which reduces, for a four-dimensional base space, to self-duality. The latter condition is met for all of the fivebranes of interest to us since they are all built on instanton gauge fields. Thus, in all cases, the essential issue is the existence of complex structures.

To count complex structures, we will use the connection between complex structures and covariantly constant spinors (a nice pedagogical discussion can be found in [35]). We start with a spinor η (in our case four-dimensional) of definite chirality ($\gamma_5 \eta = +\eta$, say) and unit normalized ($\eta^\dagger \eta = 1$). Then we define a tensor

$$J_{AB} = -i\eta^\dagger \gamma_{AB} \eta \quad (7.11)$$

which we will identify as a complex structure tensor (in tangent space indices and with indices raised and lowered by the identity metric). It is then automatic that if the spinor is covariantly constant with respect to some connection, so is J_{AB} . A simple Fierz identity argument then shows that J squares to -1 ($J_A{}^B J_B{}^C = -\delta_A{}^C$) and is indeed a complex structure.

We are now ready to construct the explicit complex structures. As was explained in the discussion after (5.22), on the fivebrane, *constant* spinors of definite four-dimensional chirality are covariantly constant. Using the Weyl representation for the four-dimensional gamma matrices, one has the following solutions of the two covariant constancy conditions:

$$\begin{aligned} \mathcal{D}_\mu(\Omega_+)\epsilon_+ &= 0 \Rightarrow \epsilon_+ = \begin{pmatrix} \chi \\ 0 \end{pmatrix} \\ \mathcal{D}_\mu(\Omega_-)\epsilon_- &= 0 \Rightarrow \epsilon_- = \begin{pmatrix} 0 \\ \chi \end{pmatrix}, \end{aligned} \quad (7.12)$$

where χ is *any* constant two-spinor (which we might as well unit normalize). Since there are three parameters needed to specify the general normalized two-spinor, there should be three independent choices for the two-spinor χ and therefore three choices for both ϵ_+ and ϵ_- . We will define the independent χ_r ($r = 1, 2, 3$) as those which give expectation values of the spin operator along the three coordinate axes:

$$\chi_r^\dagger \sigma^i \chi_r = \delta_{ir}. \quad (7.13)$$

This finally leads, with the help of (7.11), to the following set of three right- and left-handed complex structures:

$$\begin{aligned} J_1^+ &= \begin{pmatrix} i\sigma_2 & 0 \\ 0 & i\sigma_2 \end{pmatrix} & J_1^- &= \begin{pmatrix} -i\sigma_2 & 0 \\ 0 & -i\sigma_2 \end{pmatrix} \\ J_2^+ &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & J_2^- &= \begin{pmatrix} 0 & -\sigma_3 \\ \sigma_3 & 0 \end{pmatrix} \\ J_3^+ &= \begin{pmatrix} 0 & i\sigma_2 \\ i\sigma_2 & 0 \end{pmatrix} & J_3^- &= \begin{pmatrix} 0 & -\sigma_1 \\ \sigma_1 & 0 \end{pmatrix}. \end{aligned} \quad (7.14)$$

It is trivial to show that the J^+ commute with all the J^- and that they satisfy the Clifford algebra (7.9). These are precisely the conditions needed to generate (4, 4) supersymmetry in a left-right symmetric theory (or (4, 0) supersymmetry in a heterotic theory). The complex structures are thus extremely simple indeed.

Finally, we come to the questions of finiteness and need for higher-order or non-perturbative in α' corrections to our solutions. It is rather firmly established that two-dimensional nonlinear sigma models with (4, 4) supersymmetry without torsion ($B_{\mu\nu} = 0$) are in fact finite. The general proof was given quite some time ago by Alvarez-Gaume and Freedman [33] and assumes that (4, 4) supersymmetry is not explicitly broken at the quantum level. They then show that no (4, 4)-invariant counterterms - perturbative or non-perturbative- of the needed dimension can be constructed. If the theory is finite,

the beta-functions get no higher-order corrections and the choice of background fields which made the beta functions vanish at leading order must continue to make them vanish at all orders in α' . A similar result was shown in [36] to hold, relying heavily on the results of [34,27,28] for (4,4) models with torsion. (These arguments are backed up at the perturbative level by superfield non-renormalization theorems[27].) The functional form of the action must satisfy certain conditions in order to have (4,4) supersymmetry and one can see that the most general solution of these conditions corresponds precisely to our special multi-fivebrane solution.

As an aside, we mention that it has been argued that one really only needs (4,0) supersymmetry to achieve finiteness [27]. This would apply to variations on the solution described in Sect. 2 in which, for example, the gauge instanton scale size did not match the semi-wormhole throat transverse scale size or to the original instanton solution(5.24). In the discussion given earlier in this section, we recall that the existence and properties of the right-moving complex structures $f_i^{(+)}$ have nothing to do with the properties of the gauge field (which governs the left-moving complex structures). So, if we keep the same metric then we should have the same $f_i^{(+)}$ and thus at least a (4,0) supersymmetry. In this case there will be corrections to the beta functions so that the theory is not finite, but may be constructible order by order, as was shown in [18] for the solution (5.24) using spacetime methods. This subject has yet to be explored in any detail from the worldsheet point of view.

8. Algebraic CFT Approach

It is one thing to show that a sigma model is a superconformal field theory, as we have done in the previous section, and quite another to be able to classify its primary field content and calculate n-point functions of its vertex operators. Indeed, in order to answer all the interesting questions about string solitons, it would be desirable to have as detailed an algebraic understanding of the underlying conformal field theory as we already have for, say, the minimal models. We are far from having such an understanding, but in this section we will see that in some case useful progress can be made.

Recall from section 6 that the (four-dimensional part of the) metric of the symmetric solution has the form

$$ds^2 = e^{2\phi} dx^2 \quad (8.1)$$

where dx^2 is the flat metric on R^4 and

$$e^{2\phi(x)} = e^{2\phi_0} + \sum_1^n \frac{Q_i}{(x - x_i)^2} . \quad (8.2)$$

The singularities in $e^{2\phi}$ are associated with the semi-wormholes. Taking $n = 1$ and the limit $e^{2\phi_0} \rightarrow 0$ gives

$$e^{2\phi} = \frac{Q}{x^2}, \quad (8.3)$$

which is the solution corresponding to the semi-wormhole throat itself. Using spherical coordinates centered on the singularity, and defining a logarithmic radial coordinate by

$t = \sqrt{Q} \ln \sqrt{x^2/Q}$, the metric, dilaton and axion field strength of the throat may be written in the form

$$\begin{aligned} ds^2 &= dt^2 + Q d\Omega_3^2, \\ \phi &= -t/\sqrt{Q}, \\ H &= -Q\epsilon, \end{aligned} \tag{8.4}$$

where $d\Omega_3^2$ is the line element and ϵ the volume form of the unit 3-sphere obeying $\int \epsilon = 2\pi^2$. The geometry of the throat is thus a 3-sphere of radius $\sqrt{Q}$ times the open line R^1 and the dilaton is linear in the coordinate of the R^1 . Remarkably, these metric and antisymmetric tensor fields are such that the curvatures constructed from the generalized connections, defined in (5.4), are identically zero, reflecting the parallelizability of S^3 . The axion charge Q is integrally quantized. So, since Q appears in the metric, the radius of the S^3 is quantized as well.

The sigma model defined by these background fields is an interesting variant of the Wess-Zumino-Witten model and the underlying conformal field theory can, it turns out, be analyzed in complete detail. The basic observation along these lines was made in [37] in the lorentzian context and euclideanized in [38,39]: the S^3 and the antisymmetric tensor field are equivalent to the $O(3)$ Wess-Zumino-Witten model of level

$$k = \frac{Q}{\alpha'}, \tag{8.5}$$

while the R^1 and the linear dilaton define a Feigin-Fuks-like free field theory with a background charge induced by the linear dilaton. Both systems are conformal field theories of known central charges:

$$c_{wzw} = \frac{3k}{k+2} \quad c_{ff} = 1 + \frac{6}{k}. \tag{8.6}$$

The shift of the R^1 central charge away from unity is a familiar background charge effect which has been exploited in constructions of the minimal models [40] and in cosmological solutions [37].

For the combined theory to make sense, the net central charge must be four. Let us for the moment consider the bosonic string. If we expand c_{wzw} in powers of k^{-1} (this corresponds to the usual perturbative expansion in powers of α'), we see instead that

$$c_{tot} = c_{wzw} + c_{ff} = 4 + O(k^{-2} \sim \alpha'^2). \tag{8.7}$$

But, we should not have expected to do any better: the field equations we solved in section 5 to get this solution are only the leading order in α' approximation to the full bosonic string theory field equations and we must expect higher-order corrections to the fields and central charges. In fact, this issue can be studied in detail and it can be shown [41] that the metric and antisymmetric tensor fields are not modified and that the only modification of the dilaton is to adjust the background charge of the R^1 (*i.e.* the coefficient of the linear term in ϕ) so as to maintain c_{tot} exactly equal to four.

While this is quite interesting, we are really interested in the superstring case. The leading-order-in- α' metric, dilaton etc. fields are the same as in the bosonic case (and,

because of the non-renormalization theorems, we expect no corrections to them) but various fermionic terms are added to the previous purely bosonic sigma model. The structure is that of the (1,1) worldsheet supersymmetric sigma model (7.4) discussed in section 7. There is still an $S^3 \times R^1$ split, but the component theories are supersymmetrized versions of Wess-Zumino-Witten and Feigin-Fuks. The Feigin-Fuks theory is still essentially free. In the supersymmetric WZW theory, the four-fermi terms vanish identically because, as pointed out above, the generalized curvature vanishes for this background. As a consequence, the generalized connections are locally pure gauge and can be eliminated from the fermion kinetic terms by a gauge rotation of the frame field. Since the fermions are effectively free, they make a trivial addition to the central charges of both the S^3 and the R^1 models:

$$c_{wzw} = \frac{3k}{k+2} + \frac{3}{2} \quad c_{ff} = 1 + \frac{6}{k} + \frac{1}{2} . \quad (8.8)$$

There is, however, a small subtlety: the gauge rotation which decouples the fermions is *chiral*, and therefore anomalous, because the left- and right-moving fermions couple to two *different* pure gauge generalized connections, Ω_+ and Ω_- . The entire effect of this anomaly on the central charge turns out to be the replacement in c_{wzw} of k by $k-2$ (the details can be found in [42]) with the result that

$$c_{wzw} = \frac{3(k-2)}{k} + \frac{3}{2} \quad c_{tot} = c_{wzw} + c_{ff} = 6 . \quad (8.9)$$

Six is, of course, exactly the value we want for the central charge. The remarkable fact is that, in the supersymmetric theory, the expansion of c_{wzw} in powers of k^{-1} *terminates* at first non-trivial order and no modification of the dilaton field is needed to maintain the desired central charge of six. These results are consistent with the non-renormalization theorems discussed in section 7, but are not tied to perturbation theory, since they derive from exactly-solved conformal field theories. On the other hand, since the present discussion makes no reference to the (4,4) supersymmetry which was crucial in proving the perturbative non-renormalization theorems of section 7, an important element is still missing.

This is a good point to remind the reader of the hierarchy of superconformal algebras. Much of what we know about conformal field theory comes from studying the representation theory of these algebras. The basic $N=1$ superconformal algebra contains an energy-momentum tensor $T(z)$ and its superpartner $G(z)$. The essential information is contained in the algebra obeyed by their Laurent coefficients L_n and G_r :

$$\begin{aligned} [L_m, L_n] &= \frac{\hat{c}}{8} m(m^2 - 1) \delta_{m+n,0} + (m-n) L_{m+n}, \\ \{G_r, G_s\} &= \frac{\hat{c}}{2} (r^2 - \frac{1}{4}) \delta_{r+s,0} + 2L_{r+s}, \\ [L_m, G_r] &= (\frac{m}{2} - r) G_{m+r} \end{aligned} \quad (8.10)$$

with $\hat{c} = 2c/3$ in terms of the usual conformal anomaly. All superstring theories have at least this much worldsheet supersymmetry. The $N=2$ superconformal algebras differ from

this by having a conserved current $J(z)$ and *two* supercharges $G^\pm(z)$ distinguished by the value (± 1) of their charge with respect to the current $J(z)$. This charge also plays a key role in the GSO projection which rids the theory of tachyons. The important new algebraic relations are contained in the commutation relations

$$\begin{aligned} [J_m, G_r^\pm] &= \pm G_r^\pm, \\ \{G_r^+, G_s^+\} &= \{G_r^-, G_s^-\} = 0, \\ \{G_r^+, G_s^-\} &= L_{r+s} + \frac{1}{2}(r-s)J_{r+s} + \frac{\hat{c}}{4}(r^2 - \frac{1}{4})\delta_{r+s}. \end{aligned} \quad (8.11)$$

There is an $N=1$ subalgebra generated by $T(z)$ and $G(z) = \frac{1}{\sqrt{2}}(G^+(z) + G^-(z))$. The ‘practical’ utility of the $N=2$ algebra is that the conserved current defines a free field H by the relation $J(z) = i\sqrt{\frac{2}{3}}\partial_z H(z)$ and this free field can be used to construct the $N=1$ *spacetime* supersymmetry charge in a compactification to four dimensions [29]. One further extension, to four supercharges, turns out to be possible. There are now three conserved currents J^i which generate an $SU(2)$ Kac-Moody algebra and the supercharges $G^\alpha(z), \tilde{G}^\alpha(z)$ are in $I = 1/2$ representations of the conserved $SU(2)$. The relevant (anti-) commutation relations are

$$\begin{aligned} [L_m, L_n] &= (m-n)L_{m+n} + \frac{k}{2}m(m^2-1)\delta_{m+n,0}, \\ \{G_r^\alpha, G_s^\beta\} &= \{\tilde{G}_r^\alpha, \tilde{G}_s^\beta\} = 0, \\ \{G_r^\alpha, G_s^\beta\} &= 2\delta^{\alpha\beta}L_{r+s} - 2(r-s)\sigma_{\alpha\beta}^i J_{r+s}^i + \frac{k}{2}(4r^2-1)\delta_{r+s,0}, \\ [J_m^i, J_n^j] &= i\epsilon^{ijk}J_{m+n}^k + \frac{1}{2}km\delta_{m+n,0}\delta_{ij}, \\ [J_m^i, G_r^\alpha] &= -\frac{1}{2}\sigma_{\alpha\beta}^i G_{m+r}^\beta, \\ [J_m^i, \tilde{G}_r^\alpha] &= \frac{1}{2}\bar{\sigma}_{\alpha\beta}^i \tilde{G}_{m+r}^\beta \end{aligned} \quad (8.12)$$

with σ^i the usual Pauli matrices and $\bar{\sigma}^i$ their complex conjugates.

The triplet of conserved charges is what is needed to construct the larger spacetime supersymmetry algebra associated with a compactification down to six, rather than four, dimensions. The $SU(2)$ Kac-Moody algebra is of arbitrary level k , but we can see by comparison with (8.10) that the central charge c is constrained to be $6k$. Since the level is constrained by unitarity to be integer, the only allowed values of the central charge are $6, 12, \dots$. Fortunately, $c = 6$ is just what we need, and this suggests that the $N=4$ algebra will be important for us.

We will now show that a closer examination of the algebraic structure of the throat conformal field theory reveals the existence of just the right extended supersymmetry. An important clue to understanding the structure of the $(4, 4)$ superconformal symmetry comes from the fact that there must be *two* $SU(2)$ Kac-Moody symmetries: The first is part of the standard $N=4$ superalgebra. This algebra contains the energy-momentum

tensor $T(z)$, four supercurrents $G^a(z)$ and three currents $J^i(z)$ of conformal weight 1, which generate an $SU(2)$ Kac-Moody algebra of a level tied to the conformal anomaly (in our case, level one). The second is the $SU(2)$ Kac-Moody algebra of the Wess-Zumino-Witten part of the throat conformal field theory. It has a general level n , related to the area of the throat cross-section (or, equivalently, its axion charge) and is clearly distinct from the $N=4$ $SU(2)$ Kac-Moody. Since the superconformal algebra is quite tightly constrained, it is not *a priori* obvious that such an $SU(2) \otimes SU(2)$ Kac-Moody is compatible with $N=4$ supersymmetry and useful information, such as restrictions on allowed values of the central charge, might be obtained by explicitly constructing the algebra (assuming a consistent one to exist). Quite remarkably, precisely the algebra we need has already been constructed by Schoutens in [43]. A closely related version of this algebra was presented in Sevrin et. al. [44], as an alternate $N=4$ superalgebra, containing an $SU(2) \otimes SU(2) \otimes U(1)$ Kac-Moody algebra, which had been missed in previous attempts at a general classification of extended superalgebras. Similar results also appear in [45]. In what follows* we briefly summarize enough of this work to explain its significance for the throat problem and, in particular, to verify the assertions made in section 7 about the rôle of $(4,4)$ supersymmetry. In addition to establishing the presence of a $(4,4)$ superconformal symmetry, this construction is a useful starting point for studying the correspondence between the instanton moduli space and perturbations of the superconformal field theory.

The construction discovered by Sevrin et.al. goes as follows: Start with the bosonic WZW model for an $SU(2) \otimes U(1)$ group manifold (this is the geometry of the throat if we let the radius of the $U(1)$ be infinite). The conformal model contains four dimension-one Kac-Moody currents, J^i $i = 1, 2, 3$ which satisfy the Kac-Moody algebra of $SU(2)$ with level n and an additional $U(1)$ current J^0 which satisfies the $U(1)$ algebra with level one. This is supersymmetrized by adding a set of four dimension-1/2 fields ψ^a satisfying the free fermion algebra (this is motivated by the arguments given earlier in this section that the fermions in a supersymmetric WZW model are, modulo anomalies, free).

As usual, the Sugawara construction provides an energy-momentum tensor

$$T(z) = -J^0 J^0 - \frac{1}{n+2} J^i J^i - \partial \psi^a \psi^a \quad (8.13)$$

with respect to which the fields J^a (ψ^a) are primaries of weight 1 ($1/2$) and which has the expected S_{wzw} conformal anomaly

$$c_{swzw} = \frac{3n}{n+2} + 3 = 6(n+1)/(n+2) . \quad (8.14)$$

There is also a Sugawara-like construction of four real supersymmetry charges G^a , with $a = 0, \dots, 3$:

$$\begin{aligned} G^0 &= 2[J^0 \psi^0 + (1/\sqrt{n+2}) J^i \psi^i + (2/\sqrt{n+2}) \psi^1 \psi^2 \psi^3] \\ G^1 &= 2[J^0 \psi^1 + (1/\sqrt{n+2}) (-J^1 \psi^0 + J^2 \psi^3 - J^3 \psi^2) - (2/\sqrt{n+2}) \psi^0 \psi^2 \psi^3] \end{aligned} \quad (8.15)$$

* This discussion was developed in collaboration with E. Martinec.

(plus cyclic expressions for G^2 and G^3). These supercharges could have been packaged as a complex $I = 1/2$ multiplet, as in (8.12). The operator product expansion of these supercharges with themselves reads

$$\begin{aligned} G^a(z)G^b(w) = & 4\frac{(n+1)}{(n+2)}\delta^{ab}(z-w)^{-3} + 2\delta^{ab}T(w)(z-w)^{-1} \\ & -8\left[\frac{1}{n+2}\alpha_{ab}^{+i}A_i^+(w) + \frac{n+1}{n+2}\alpha_{ab}^{-i}A_i^-(w)\right](z-w)^{-2} \\ & -4\left[\frac{1}{n+2}\alpha_{ab}^{+i}\partial A_i^+(w) + \frac{n+1}{n+2}\alpha_{ab}^{-i}\partial A_i^-(w)\right](z-w)^{-1} \end{aligned} \quad (8.16)$$

where

$$\alpha_{ab}^{\pm i} = \pm \delta_{[a}^i \delta_{b]}^0 + \frac{1}{2}\epsilon_{iab} \quad (8.17)$$

and

$$A_i^- = \psi^0\psi^i + \epsilon_{ijk}\psi^j\psi^k \quad A_i^+ = -\psi^0\psi^i + \epsilon_{ijk}\psi^j\psi^k + J^i \quad (8.18)$$

are commuting $SU(2)$ Kac-Moody algebras of levels 1 and $n+1$, respectively. The c -number term (the central charge) and the term involving $T(z)$ are obligatory in any higher- N superalgebra, while the terms involving dimension 1 operators are what differentiate the various possible extended superalgebras. With further effort, one shows that the $G \cdot A^\pm$ OPE generates combinations of G^a and ψ^a while the $G \cdot \psi$ OPE yields $A^\pm$ and J^0 . No new operators appear in further iterations, so the complete algebra generated by the supercharges contains just T (dimension 2), G^a (dimension 3/2), $A_\pm^i$ and J^0 (dimension 1) and ψ^a (dimension 1/2). The Kac-Moody algebra defined by the dimension 1 operators is evidently $SU(2) \times SU(2) \times U(1)$, which accords with our expectations derived from the throat geometry.

The superalgebra whose construction we have outlined above is a particular example of a one-parameter family of $N=4$ algebras dubbed the A_γ algebras. The only problem with it is that the sigma model analysis of extended supersymmetry (see for example [28]) makes quite clear that the canonically defined supercharges and energy-momentum tensor must satisfy the standard $N=4$ algebra, which closes on T , G^a and a single level-one $SU(2)$ Kac-Moody algebra J^i . The supercharges defined above obviously do not have that property. However, if we 'improve' them as follows

$$\tilde{T} = T - \frac{1}{\sqrt{n+2}}\partial J^0 \quad \tilde{G}^a = G^a - \frac{1}{\sqrt{n+2}}\partial\psi^a, \quad (8.19)$$

we can show that $\tilde{T}$, $\tilde{G}^a$ and $A_i^\pm$ (the level-one Kac-Moody current) close on themselves and enjoy precisely the standard $N=4$ superalgebra. This says that the full algebra has the standard algebra as a subalgebra, perhaps no great surprise.

This improvement has a simple physical interpretation: J^0 generates a $U(1)$ symmetry which can be regarded as a translation in a free coordinate ρ (that is, we can write $J(z) \sim \partial_z \rho(z)$ where ρ is a free scalar field). The original algebra (8.16) makes no reference to the dilaton and corresponds physically to a constant dilaton field. It is well-known

that, if one turns on a dilaton which is *linear* in a free coordinate ρ , this has the effect of adding a term proportional to $\partial_z^2 \phi \sim \partial_z J^0(z)$ to $T(z)$ and shifting the central charge of the superconformal algebra by a constant. With a little care we can show that the linear dilaton implicit in (8.19) is precisely what we obtained earlier in this section in our discussion of the WZW-Feigin-Fuks conformal field theory of the throat. This is a further piece of evidence that the improved energy-momentum tensor $\tilde{T}$ is the physically relevant one. Now comes the miracle: T is, in any event, not physically acceptable because it has a central charge of $6(n+1)/(n+2)$. The central charge of $\tilde{T}$, however, can easily be shown to be 6, precisely the required value! This same construction has also appeared in a recent study by Kounnas *et al.* [46] of Liouville theories with $N = 4$ extended supersymmetry.

This shows that there is an exact conformal field theory of just the right central charge associated with the throat geometry and verifies the key rôle of $N=4$ extended supersymmetry in establishing the physics of the model. There are many fivebrane applications of this exact throat conformal field theory which are just beginning to be worked out. Perhaps the most interesting concern the vertex operators of excitations about the wormhole throat, among which one must find the moduli of the exact solutions. In any event, this line of argument shows that the dramatic consequences of (4,4) superconformal symmetry, which we first extracted from perturbative considerations, seem to have nonperturbative status.

9. Other Supersymmetric Solitons

For completeness we will briefly mention in this section some of the other supersymmetric string solitons which have been discussed in the literature. These solitons are mainly understood from a spacetime point of view; a worldsheet analysis of the type discussed in this paper remains to be performed.

In [47,48], supersymmetric one-brane solutions were found. The construction of this solution involves an intriguing interplay between spacetime and worldsheet methods. These “solutions” do not solve the equations of motion following from (5.1) in the usual sense, rather there is a singularity at the origin corresponding to the presence of a zero-thickness fundamental string. The form of this singularity is then precisely dictated by the known coupling of spacetime vertex operators to the string worldsheet.

In [5] a one-brane solution of heterotic string theory was found which is an everywhere smooth solution of the equations of motion of (5.1). The construction of this solution involves crucially the properties of octonions. One of the many bizarre features of this soliton is that it breaks 15 of the 16 supersymmetries of ten-dimensional Minkowski space, in contrast to previously known examples of supersymmetric solitons which all break half of the supersymmetries. Clearly this is the odd duck in the family of supersymmetric string solitons. The fact that this solution does not have finite energy per unit length further complicates its physical implications.

The fivebrane solutions discussed in these lectures are characterized by a nonvanishing integral of the three-form field strength H , while the abovementioned string solutions carry a non-zero integral of the dual of H . In type II strings, there are a variety of d -form field strengths. These fields are not present in the heterotic string because they are created

by Ramond-Ramond vertex operators. For each field there exists [25] a p -brane solution carrying the corresponding charge or its dual. A conformal field theoretic description of these solutions is a challenging and interesting problem because of the non-trivial Ramond-Ramond backgrounds.

The possibility has often been contemplated that our universe is in fact a three-brane embedded in a higher-dimensional space. The difficulty with this idea is the opposite of the difficulty associated with Kaluza-Klein compactification: there are too few solutions rather than too many, and none with the right properties have been found. In particular, although it is easy to obtain scalar and spin $1/2$ zero modes on such a membrane, it has not been possible until recently to find theories with higher spin zero modes. In [18] it was shown that there do exist five-brane solutions of type II string theory which possess not only spin zero and spin $1/2$ zero modes but also spin one zero modes. Unfortunately, the construction presented there also suggests that spin two zero modes cannot be found by the same mechanism. As a perhaps more realistic example, the p -brane solutions of [25] do in fact contain a three-brane solution[49]. It is associated with the self-dual five-form of the chiral IIB theory. The effective four-dimensional theory has extended $N = 4$ supersymmetry. This may be a useful model for exploring this alternate method of getting rid of extra dimensions.

Finally there is the issue of solitons in the context of string compactification to four dimensions[17,50]. This is a rich subject—too rich to review here. One hope is that ‘realistic’ string solitons may have peculiar enough properties that their detection could serve as a ‘smoking gun’ for string theory. For example in [51] it is argued that the fractional charge carried by certain string solitons could serve this purpose. Of course direct detection of string solitons is generally difficult because they are typically ultramassive, but perhaps there are surprises in store.

10. Conclusion

In these lectures, we have constructed a special set of conformal field theories which have the interpretation of soliton solutions of heterotic string theory. We first constructed them as solutions of the leading order in α' beta function conditions and then showed that, owing to an extended worldsheet supersymmetry, the associated nonlinear sigma model is an exact conformal field theory. It is the existence of an explicit and exact conformal field theory associated with the soliton solution which distinguishes the solution described here from previous attempts to construct string theory solitons. There are several lines of inquiry which can be pursued now that “exact” string solitons exist. One issue concerns the mass of the soliton. In all previous discussions of string solitons, the mass has been computed using the lowest-order spacetime effective action (5.1). It would obviously be desirable to have a purely conformal field theoretic definition of the mass—perhaps in terms of some correlation functions. Our exact soliton conformal field theory should provide a useful context for addressing this question. A second issue is the question of stringy collective coordinates and their semiclassical quantization. It should be an instructive challenge to translate the well-known standard field theory physics of collective coordinates into the string theory context. This is a nontrivial matter because motion in

collective coordinate space becomes motion in a space of conformal field theories and it is a nontrivial matter to find the action associated with such motions (and knowing the exact underlying conformal field theories should help). Yet another question to pursue is that of stringy black hole physics. We noted that our solitons were similar to the extreme Reissner-Nordstrom black holes in the sense that, while they have no singularity or event horizon, if one increases their mass by any finite amount (while keeping the axion charge fixed), an event horizon and a singularity (lying at a finite geodesic distance from any finite point) will appear. Such black hole solitons can easily be created by scattering some external particle on an extremal soliton and, by studying stringy scattering theory about the extremal soliton, one should be able to explore, in a controlled way, how a stringy black hole Hawking radiates and the nature of the final state it approaches. These are quite difficult questions, but having precise control of the underlying conformal field theory may allow us to make progress on them. Perhaps it will be possible to report on progress along these lines at the next Trieste School.

Acknowledgements

We would like to thank Emil Martinec for collaboration on the material contained in section 8. This work was supported in part by DOE grants DE-AC02-84-1553 and 8-48061-2500-3, and by NSF grant PHY90-00386. J.H. also acknowledges the support of a NSF PYI award PHY91-96117. Some of the material in these lectures was presented by C.G.C. at the Sixth J. A. Swieca School of Theoretical Physics (January 1991) [52].

References

- [1] E. Witten, *On string theory and black holes*, Phys. Rev. **D44** (1991) 314.
- [2] A. Strominger, *Heterotic Solitons*, Nuc. Phys. **B343** (1990) 167.
- [3] See e.g. R. Rajaraman, *Solitons and Instantons*, North-Holland, 1982.
- [4] E. Witten and D. Olive, *Supersymmetry Algebras that include Topological Charges*, Phys. Lett. **78B** (1978) 97.
- [5] J. Harvey and A. Strominger, *Octonionic Superstring Solitons*, Phys. Rev. Lett. **66** (1991) 549.
- [6] S. Coleman, *The Uses of Instantons in Aspects of Symmetry*, Cambridge University Press, 1985.
- [7] M.F. Atiyah, N. J. Hitchin, and I. M. Singer, *Deformations of Instantons*, Proc. Natl. Acad. Sci. USA **74** (1977) 2662.
- [8] M. F. Atiyah, V. G. Drinfeld, N. J. Hitchin and Y. I. Manin, *Construction of Instantons*, Phys. Lett. **65A** (1978) 185.
- [9] B. Zumino, Phys. Lett. **B69**, (1977) 369.
- [10] M. K. Prasad and C. M. Sommerfield, *Exact Classical Solutions for the 't Hooft Monopole and Julia-Zee Dyon*, Phys. Rev. Lett. **35** (1975) 760.
- [11] E. Corrigan and P. Goddard, *An n -Monopole Solution with $(4n-1)$ Degrees of Freedom*, Commun. Math. Phys. **80** (1981) 575.
- [12] N. S. Manton, *The Force Between 't Hooft-Polyakov Monopoles*, Nucl. Phys. **B126** (1977) 525.
- [13] C. Montonen and D. Olive, *Magnetic Monopoles as Gauge Particles?*, Phys. Lett. **72B** (1977) 117.
- [14] P. Goddard, J. Nuyts and D. Olive, Nucl. Phys. **B125** (1977) 1.
- [15] H. Osborn, *Topological Charges for $N = 4$ Supersymmetric Gauge Theories and Monopoles of Spin 1*, Phys. Lett. **83B** (1979) 321.
- [16] M. Green and J. Schwarz, Phys. Lett. **B151** (1985) 21.
- [17] E. Witten, *Topological Tools in Ten-Dimensional Physics in Unified String Theories*, World-Scientific, 1986.
- [18] C. Callan, J. Harvey and A. Strominger, *Worldbrane Actions for String Solitons*, preprint PUPT-1244, EFI-91-12 (1991).
- [19] M. J. Duff and J. X. Lu, *Elementary Fivebrane Solutions of $D=10$ Supergravity*, Nucl. Phys. **B354** (1991) 141.
- [20] M. J. Duff and J. X. Lu, *Remarks on String/Fivebrane Duality*, Nucl. Phys. **B354** (1991) 129.
- [21] M. J. Duff and J. X. Lu, *Strings from Fivebranes*, Phys. Rev. Lett. **66** (1991) 1402.
- [22] A. Strominger, *Superstrings with Torsion*, Nuclear Physics **B274**, 253 (1986).

- [23] E. A. Bergshoeff and M. de Roo, *The Quartic Effective Action of the Heterotic Superstring and Supersymmetry*, Nuc. Phys. **B328** (1989) 439.
- [24] E. Witten, Comm. Math. Phys. **92** (1984) 455.
- [25] G. Horowitz and A. Strominger, *Black strings and p-branes*, Nucl. Phys. **B360** (91) 197.
- [26] S.B. Giddings and A. Strominger, *Exact black fivebranes in critical superstring theory*, Phys. Rev. Lett. **67** (1991) 2930.
- [27] P. S. Howe and G. Papadopoulos, *Finiteness of $(4,q)$ Models with Torsion*, Nucl. Phys. **B289** (1987) 264; *Further Remarks on the Geometry of Two-Dimensional Non-linear σ Models*, Class. Quant. Grav. **5** (1988) 1647.
- [28] P. van Nieuwenhuizen and de Wit, *Rigidly and Locally Supersymmetric Two-Dimensional Nonlinear Sigma Models with Torsion*, Nucl. Phys. **B312** (1989) 58.
- [29] T. Banks, L. J. Dixon, D. Friedan and E. Martinec, *Phenomenology and Conformal Field Theory or, Can String Theory Predict the Weak Mixing Angle?*, Nucl. Phys. **B299** (1988) 613.
- [30] T. Banks and L. Dixon, *Constraints on String Vacua with Space-time Supersymmetry*, Nucl. Phys. **B307** (1988) 93.
- [31] A review by C. Callan and L. Thorlacius can be found in *Particles, Strings and Supernovae*, Vol. 2, eds. A. Jevicki and C.-I. Tan, World Scientific (1989).
- [32] C. Callan, D. Friedan, E. Martinec and M. Perry, *Strings in Background Fields*, Nuc. Phys. **B262** (1985) 593.
- [33] L. Alvarez-Gaume and D. Z. Freedman, *Geometrical Structure and Ultraviolet Finiteness in the Supersymmetric Sigma Model*, Comm. Math. Phys. **80** 443 (1981).
- [34] S. J. Gates, C. M. Hull and M. Rocek, *Twisted Multiplets and New Supersymmetric Nonlinear Sigma Models*, Nucl. Phys. **B248** (1984) 157.
- [35] P. Candelas, *Lectures on Complex Manifolds in Superstrings and Grand Unification, Proceedings of the Winter School on High Energy Physics, Puri, India, 1988*, T. Pradhan ed., World Scientific Publishers.
- [36] C. Callan, J. Harvey and A. Strominger, *Worldsheet approach to heterotic solitons and instantons*, Nucl. Phys. **B359** (1991) 611.
- [37] I. Antoniadis, C. Bachas, J. Ellis and D. Nanopoulos, *Cosmological String Theories and Discrete Inflation*, Phys. Lett. **B211** (1988) 393 and Nucl. Phys. **B328** (1989) 117.
- [38] G. Horowitz, A. Strominger and S. J. Rey, unpublished.
- [39] S. J. Rey, *Confining Phase of Superstrings and Axionic Strings*, Phys. Rev. **D43** (1991) 439.
- [40] V. Dotsenko and V. Fateev, *Conformal Algebra and Multipoint Correlation Functions in Two-Dimensional Statistical Models*, Nucl. Phys. **B240** (1984) 312.

- [41] R. Khuri, *Some Instanton Solutions in String Theory*, Princeton preprint PUPT-1219 (1990).
- [42] R. W. Allen, I. Jack, D. R. T. Jones, *Chiral σ Models and the Dilaton β Function*, Z. Phys. C **41** (1988) 323 and references therein.
- [43] K. Schoutens, Nucl. Phys. **B295** [FS21] (1988) 634.
- [44] A. Sevrin, W. Troost and A. van Proeyen, *Superconformal Algebras in Two Dimensions with $N=4$* , Phys. Lett. **B208** (1988) 447.
- [45] E. Ivanov, S. Krivinos and V. Levant, Nucl. Phys. **B304** (1988) 601 ;
Phys. Lett. **B215** (1989) 689.
- [46] C. Kounnas, M. Porrati and B. Rostand, Phys. Lett. **B258** (1991) 61.
- [47] A. Dabholkar and J. Harvey, *Nonrenormalization of the Superstring Tension*, Phys. Rev. Lett. **63** (1989) 719.
- [48] A. Dabholkar, G. Gibbons, J. Harvey and F. R. Ruiz, *Superstrings and Solitons*, Nucl. Phys. **B340** (1990) 33.
- [49] M. J. Duff and J. X. Lu, *The Self-Dual Type II-B Superthreebrane*, CTP-TAMU-29/91 (to appear in Phys. Lett. B).
- [50] T. Banks, M. Dine, H. Dijkstra and W. Fischler, *Magnetic Monopole Solutions of String Theory*, Phys. Lett. **B212** (1988) ;
B. Greene, A. Shapere, C. Vafa, and S. Yau, *Stringy Cosmic Strings and Noncompact Calabi-Yau Manifolds*, Nucl. Phys. **B337** (1990) 1 ;
M. Cvetič, F. Quevedo and S. J. Rey, *Stringy Domain Walls and Target Space Modular Invariance*, Phys. Rev. Lett. **67** (1991) 1836 ;
J. Harvey and J. Liu *Magnetic Monopoles in $N=4$ Supersymmetric Low-Energy Superstring Theory* EFI-90-82, PUPT-1259 (1991).
- [51] X. G. Wen and E. Witten, *Electric and Magnetic Charges in String Theory*, Nucl. Phys. **B261** (1985) 651 ;
G. Athanasiu, J. Atick, M. Dine and W. Fischler, *Remarks on Wilson Lines, Modular Invariance and Possible String Relics in Calabi-Yau Compactification*, Phys. Lett. **B214** (1988) 651.
- [52] Proceedings of the Sixth J. A. Swieca School of Theoretical Physics, O. Eboli ed.

INTRODUCTION TO W -ALGEBRAS

Adel Bilal

Theory Division, CERN, 1211 Geneva 23, Switzerland

and

*Laboratoire de Physique Théorique de l'Ecole Normale Supérieure,
24 rue Lhomond, 75231 Paris Cedex 05, France*

ABSTRACT

The subject of W -algebras is by now rather large, and it encompasses many areas of theoretical physics. I cannot describe them all in detail. Therefore, I have chosen to present in the first lecture, after a brief introduction to what W -algebras are, an (incomplete) overview of various domains where W -algebras appear.

The second and third lectures contain a rather detailed discussion of the free-field realization of W -algebras: construction of the W -generators, highest-weight representations, the role of the screening operators, null vectors and degenerate representations, operator product algebra and fusion rules, minimal models, unitary models and their coset construction.

The fourth lecture dealt with classical W -algebras. It is not described in these notes, and the reader is referred to a recent preprint.

1. Lecture : Introduction

1.1 WHAT ARE W -ALGEBRAS ?

W -algebras first appeared in 2D conformal field theory. The standard example is Zamolodchikov's W_3 algebra [1]. It contains two infinite sets of generators, L_m and W_m , $m \in \mathbb{Z}$. The L_m 's just satisfy a Virasoro subalgebra with central charge c :

$$[L_n, L_m] = (n - m)L_{n+m} + \frac{c}{12}(n^3 - n)\delta_{n+m,0} . \quad (1.1)$$

The commutator of the L_n with the W_m

$$[L_n, W_m] = (3n - (n + m))W_{n+m} \quad (1.2)$$

tells us that the field $W(z) = \sum_n W_n z^{-n-3}$ is a primary field of conformal dimension (spin) 3. So far there is nothing special. The interesting commutator is that of two W 's:

$$\begin{aligned} [W_n, W_m] = & \frac{c}{3 \cdot 5!}(n^2 - 4)(n^2 - 1)n\delta_{n+m,0} \\ & + (n - m) \left[\frac{1}{15}(n + m + 2)(n + m + 3) - \frac{1}{6}(n + 2)(m + 2) \right] L_{n+m} \\ & + \beta^2(n - m)\Lambda_{n+m} \end{aligned} \quad (1.3)$$

where

$$\beta^2 = \frac{16}{5c + 22} , \quad \Lambda_n = \sum_k : L_k L_{n-k} : + \frac{1}{5}x_n L_n \quad (1.4)$$

where $x_{2n} = (1 + n)(1 - n)$ and $x_{2n+1} = (2 + n)(1 - n)$.

One may take two attitudes. Either define a new set of generators Λ_n , and consider the commutators of this Λ_n with L_m and W_m . This introduces still other generators, and one has to define fields of higher and higher spin. This never stops. The other attitude is to accept the non-linear structure of the algebra (it is thus not a Lie algebra) and keep $\Lambda(z) =: T(z)T(z) : - \frac{3}{10}\partial^2 T(z)$ as a composite field. Closure is then obtained in the enveloping algebra of the L_m 's and W_m 's. This second approach has proved more fruitful.

For completeness let me give the operator product expansions (OPEs) equivalent to the above commutation relations:

$$\begin{aligned}
 T(z)T(z') &= \frac{c/2}{(z-z')^4} + \frac{2T(z')}{(z-z')^2} + \frac{\partial T(z')}{z-z'} + \left[\Lambda(z') + \frac{3}{10} \partial^2 T(z') \right] \\
 T(z)W(z') &= \frac{3W(z')}{(z-z')^2} + \frac{\partial W(z')}{z-z'} \\
 W(z)W(z') &= \frac{c/3}{(z-z')^6} + \frac{2T(z')}{(z-z')^4} + \frac{\partial T(z')}{(z-z')^3} \\
 &\quad + \frac{3}{10} \frac{\partial^2 T(z')}{(z-z')^2} + \frac{1}{15} \frac{\partial^3 T(z')}{z-z'} + \frac{2\beta\mathcal{L}(z')}{(z-z')^2} + \frac{\beta\partial\mathcal{L}(z')}{z-z'}
 \end{aligned} \tag{1.5}$$

where the terms of order $(z-z')^0$ in the expansion of T with itself define the field Λ .

In general, a W_n -algebra contains generators L_m , $W_m^{(3)}$, $W_m^{(4)}$, ..., $W_m^{(n)}$ corresponding to the energy-momentum tensor and primary fields of conformal dimensions (spins) 3, 4, ..., n with OPEs

$$\begin{aligned}
 T(z)W^{(k)}(z') &= \frac{kW^{(k)}(z')}{(z-z')^2} + \frac{\partial W^{(k)}(z')}{z-z'} \\
 W^{(k)}(z)W^{(l)}(z') &= \sum_{r=0}^{k+l-1} : \frac{\text{differential polynomial } (W^{(s)})(z')}{(z-z')^{k+l-r}} : .
 \end{aligned} \tag{1.6}$$

As usual the mode expansion for a spin k field is $W^{(k)}(z) = \sum_n W_n^{(k)} z^{-n-k}$.

These W_n -algebras have series of unitary representations for central charges

$$c = (n-1) \left(1 - \frac{n(n+1)}{p(p+1)} \right), \quad p = n+1, n+2, \dots \tag{1.7}$$

Notice that for $n=2$ this is nothing but the FQS series $c = 1 - 6/p(p+1)$.

1.2 RELATIONS WITH OTHER AREAS OF THEORETICAL PHYSICS

There are many relations with other areas of theoretical/mathematical physics which provide many different motivations to study W -algebras.

2D conformal field theory

Conformal field theories, and in particular those with additional W -symmetries, describe critical points of 2D statistical mechanics models, e.g. the W_3 -algebra with $c = 4/5$ describes the 3-state Potts model. In general, W_n -algebras describe statistical mechanical models with some Z_n -symmetry. There is an interesting observation [2] about the 3-state Potts model. It is also contained in the standard FQS unitary series ($p = 5$), but corresponds to an exceptional non-diagonal modular invariant:

$$Z = |\chi_0 + \chi_3|^2 + |\chi_{2/5} + \chi_{7/5}|^2 + 2|\chi_{2/3}|^2 + 2|\chi_{1/15}|^2 \quad (1.8)$$

where χ_h denotes the character of the Virasoro representation with highest weight h . One observes that $h = 3$ appears combined with the trivial representation. This is a sign of a bigger symmetry generated by a spin 3 field. Indeed, in terms of characters of the W_3 -algebra, the partition function is rewritten as

$$Z = |\hat{\chi}_0|^2 + |\hat{\chi}_{2/5}|^2 + |\hat{\chi}_{2/3,+}|^2 + |\hat{\chi}_{2/3,-}|^2 + |\hat{\chi}_{1/15,+}|^2 + |\hat{\chi}_{1/15,-}|^2. \quad (1.9)$$

Clearly, when we consider the bigger W_3 -algebra instead of the Virasoro algebra alone, $W_{-3}|0\rangle$ and $|0\rangle$ belong to one and the same representation. When written in terms of the characters of W_3 , the partition function is diagonal. This is a general feature; the bigger the algebra with respect to which one writes down the characters, the “more diagonal” is the partition function, and the more one can say about a model solely from representation theory. This is an obvious motivation to find the biggest (chiral) symmetry algebra of a given model and study its representations.

Affine Lie algebras

Recall that the affine Lie algebra $\hat{su}(n)_k$ is defined by the commutation relations

$$[J_l^a, J_m^b] = f^{abc} J_{l+m}^c + k l \delta_{l+m,0} \delta^{ab} \quad (1.10)$$

where the f^{abc} are the structure constants of $su(n)$, and k is an integer called the

level (provided we have normalized the roots to have squared length equal to 2). Consider then the coset

$$\frac{\hat{su}(n)_k \otimes \hat{su}(n)_1}{\hat{su}(n)_{k+1}} \quad (1.11)$$

where the quotient is taken with respect to the diagonal subalgebra. This coset provides a unitary representation of the W_n -algebra [2] with central charge

$$c = (n-1) \left(1 - \frac{n(n+1)}{(k+n)(k+1+n)} \right). \quad (1.12)$$

The energy-momentum tensor is given by the coset Sugawara construction $T = T_k + T_1 - T_{k+1}$ with $T_k \sim J^a J^a$: etc. Similarly, the $W^{(l)}$, $l = 3, \dots, n$, are obtained as $W^{(l)} \sim J \dots J$ with l factors of the currents J corresponding to the Casimir invariant of order l . A sort of field theoretic version of these coset models is provided by the gauged WZWN-models [3].

Integrable models of the KdV-type in 1+1 dimensions (KP-hierarchy)

Consider a differential operator of order n in one variable σ

$$\mathcal{D}_n(\sigma) \equiv \partial_\sigma^n - \sum_{k=2}^n u_k(\sigma) \partial_\sigma^{n-k} \quad (1.13)$$

associated with $sl(n) \sim su(n)$ (again, there is one coefficient function u_k for any Casimir invariant C_k of these algebras). Introduce the algebra of pseudo-differential operators, generated by powers of ∂ and its inverse ∂^{-1} . In this algebra it is possible to define $\mathcal{D}_n^{1/n} = \partial - \frac{1}{n} u_2 \partial^{-1} + \dots$ such that $(\mathcal{D}_n^{1/n})^n = \mathcal{D}_n$. Suppose now that the functions u_k depend on some extra "time" t_r ($r \in \mathbb{Z}_+$), and define a time evolution by

$$\frac{\partial \mathcal{D}_n}{\partial t_r} = \left[\left(\mathcal{D}_n^{r/n} \right)_+, \mathcal{D}_n \right] \quad (1.14)$$

where the subscript $+$ means to take only the differential operator part (positive or zero powers of ∂). To fix the ideas, consider the example $n = 2$, $r = 3$, where $\mathcal{D}_2 = \partial_\sigma^2 - u(\sigma)$. Then $\mathcal{D}_2^{1/2} = \partial - \frac{u}{2} \partial^{-1} + \frac{u'}{4} \partial^{-2} + \dots$ and $(\mathcal{D}_2^{3/2})_+ = \partial^3 - \frac{3}{2} u \partial - \frac{3}{4} u'$.

One then easily checks that $\left[\left(\mathcal{D}_2^{3/2} \right)_+, \mathcal{D}_2 \right] = \frac{3}{4} u u' - \frac{1}{4} u'''$ which by the evolution

equation is set equal to $-\dot{u} \equiv -\partial u / \partial t$ where $t \equiv t_3$. Upon rescaling $t \rightarrow -4t$ and $u \rightarrow 2u$ this is the standard form of the KdV equation:

$$\dot{u} = -u''' + 6uu' . \quad (1.15)$$

For general n , the evolution equations yield the generalized KdV equations (reductions of the KP-hierarchy) of the form $\frac{\partial u_k}{\partial t_r} = f_{k,r}(u_l, u'_l, \dots)$. The interesting thing about these evolution equations is that one can define a Poisson bracket $\{\cdot, \cdot\}$ between the u_k 's (the Gelfand-Dikii bracket) [4,5] and Hamiltonians H_r , such that

$$\frac{\partial \mathcal{D}_n}{\partial t_r} = \{\mathcal{D}_n, H_r\} , \quad H_r = \int d\sigma F_r(u_k, u'_k \dots) . \quad (1.16)$$

To illustrate this, consider again the KdV equation where we define the Poisson bracket

$$\{u(\sigma), u(\sigma')\} = \frac{6\pi}{c} (-\partial_\sigma^3 + 4u(\sigma)\partial_\sigma + 2u'(\sigma)) \delta(\sigma - \sigma') \quad (1.17)$$

and the Hamiltonian $H = \frac{c}{6\pi} \int u^2$. Then, upon expanding into Fourier modes $u(\sigma) = \frac{6}{c} \sum_r l_r e^{-ir\sigma} - \frac{1}{4}$ one obtains

$$\{l_r, l_s\} = (r-s)l_{r+s} + \frac{c}{12}(r^3 - r)\delta_{r+s,0} \quad (1.18)$$

which is nothing but the Virasoro algebra realized by Poisson brackets [6]. Equivalently the Poisson bracket $\{u(\sigma), u(\sigma')\}$ corresponds to the OPE of T with itself. In general, the Gelfand-Dikii Poisson bracket $\{u_k(\sigma), u_l(\sigma')\}$ corresponds to the OPEs of the W_n -algebra [7] (see also [8]). In fact it is precisely equivalent to the one-contraction parts of the OPEs, thus neglecting normal ordering effects which are always disregarded when computing Poisson brackets.

Other integrable theories in 1 + 1 dimensions: Toda field theories

Toda field theories are field theories of $n-1$ interacting bosonic fields φ_i with equations of motion

$$\partial_+ \partial_- \varphi_i = \exp \left(\sum_{j=1}^{n-1} K_{ij} \varphi_j \right) \quad (1.19)$$

where x_+, x_- are light-cone coordinates and K is the $(n-1) \times (n-1)$ Cartan matrix of $A_{n-1} \sim su(n)$: $K_{ij} = 2\delta_{i,j} - \delta_{i,j-1} - \delta_{i,j+1}$. These equations of motion can be

solved by some special ansatz expressing the φ_i in terms of $n - 1$ free left-moving and $n - 1$ free right-moving fields. It can then be shown that $\exp(-\varphi_1(x_+, x_-))$ satisfies two ordinary differential equations of order n , one in x_- and one in x_+ :

$$\mathcal{D}_n(x_-)e^{-\varphi_1(x_+, x_-)} = \tilde{\mathcal{D}}_n(x_+)e^{-\varphi_1(x_+, x_-)} = 0 \quad (1.20)$$

where the differential operators $\mathcal{D}_n(x_-)$ and $\tilde{\mathcal{D}}_n(x_+)$ are of the type discussed above. The nice thing about the Toda theories [9] is that the canonical Poisson bracket derived from the Toda action induces a Poisson bracket between the coefficient functions u_k of $\mathcal{D}_n(x_-)$ (and similarly for the $\tilde{u}_k$ in $\tilde{\mathcal{D}}_n(x_+)$) that is precisely the Gelfand-Dikii Poisson bracket discussed above leading to the W_n -algebra. Thus Toda theories provide a field theoretic realization of the Gelfand-Dikii bracket.

Integrable lattice models in 2D : Interaction Round a Face (IRF) models

Consider a statistical mechanical system defined on a square lattice with some fluctuation variables $\sigma_a, \sigma_b, \dots$ placed on the lattice sites. The probability of a given configuration is defined to be the product over all squares (faces) F_i of Boltzmann weights B assigned to every plaquette

$$\text{Prob}(F1, F2, \dots) = B(F1)B(F2)B(F3) \dots \quad (1.21)$$

Each Boltzmann weight depends on the fluctuation variables on the four lattice sites around the face (hence the name IRF), e.g.

$$B(F1) = B \left(\begin{array}{cc} \sigma_a & \sigma_b \\ \sigma_d & \sigma_c \end{array} \middle| u, p \right) \quad (1.22)$$

where u is a spectral parameter and p is a temperature-type parameter measuring the distance from the critical point. The fluctuation variables σ_a are essentially labelled by highest weights of $A_{n-1} \sim su(n)$. The Boltzmann weights are explicitly given as ratios of theta-functions [10]. For $su(2)$ the face models are nothing but the well-known hight or SOS models. The IRF models exhibit phase transitions at $p \rightarrow 0$ where the theta functions become trigonometric. One can compute the corresponding exponents. They have been found to coincide with the spectrum of conformal dimensions of the W_n -theories for appropriate central charge (determined by another free parameter of the IRF model). Furthermore, the Boltzmann weights themselves become (in the limit of infinite spectral parameter) the braiding matrices B_{dc}^{ab} of the W_n -algebra conformal field theories [11].

Random multi-matrix models

The partition function Z of these multi-random-matrix models is conjectured to be the square of the τ -function of the KP-hierarchy. From this property have been derived a set of constraints on $Z^{\frac{1}{2}}$ which can be written as

$$W_m^{(k)} Z^{\frac{1}{2}} = 0 \quad (1.23)$$

where $m \geq 1-k$ and $k = 2, 3, \dots, n$. The $W_m^{(k)}$ can be shown to satisfy a W_n -algebra [12].

Other issues

So far we have discussed W -algebras associated to the Lie algebras $A_{n-1} \sim su(n)$. There exist also W -algebras for other Lie algebras such as $B_n, C_n, D_n, \dots$

There are the so-called $W_\infty, w_\infty, W_{1+\infty}, W_\infty(\lambda)$ -algebras [13]. These include higher-spin generators with arbitrarily high spin (e.g. $k = 2, 3, \dots$ for W_∞). They are again linear (Lie) algebras. In a sense, this can be understood, since every time a composite field appears, like the Λ -field for W_3 , it can be hidden in some new generator of higher spin. Although these W_∞ -algebras are interesting in their own right, they do not share the non-linear structure of the W_n -algebras which is an essential property, and I will not discuss them here further.

Super W -algebras: The so-called minimal super W_n -algebras which are supersymmetric extensions of the minimal W_n -theories exist for central charges

$$c_n = (n-1) \left(1 - \frac{n(n+1)}{2n(2n+1)} \right) \quad (1.24)$$

i.e. these are the W_n -theories with $p = 2n$ corresponding to the coset

$$\frac{\hat{su}(n)_1 \otimes \hat{su}(n)_n}{\hat{su}(n)_{1+n}}. \quad (1.25)$$

Let us note that for $n = 3$ one has $c = \frac{10}{7}$ which appears also in the standard superconformal series ($m = 12$), where there exists an exceptional non-diagonal modular invariant Z_{E_6, D_8} . This modular invariant exhibits a bigger symmetry (super W_3 -symmetry) and becomes diagonal when written in terms of the characters of the bigger super W_3 -algebra [14] much in the same way as was discussed for the 3-state Potts model above.

2. and 3. Lectures : The free-field approach to W_n -algebras

In this and in the next lecture we are going to describe the free-field approach or Feigin-Fuchs type [15] construction of W_n -algebras. These two lectures are mainly based on refs. 1 and 16. For further details the reader might also wish to consult ref. 17.

2.1 THE CONSTRUCTION OF THE GENERATORS

One wants to construct $n - 1$ fields $T, W^{(3)}, W^{(4)}, \dots, W^{(n)}$ out of free fields. To do so one clearly needs $n - 1$ such free fields. The Hilbert space of these $n - 1$ free fields will eventually turn out to be bigger than the representation space of the W_n -algebra due to certain symmetries and also due to the presence of null vectors. This is a necessary complication, and using less than $n - 1$ free fields obviously yields a much too small Hilbert space where one can only get less than $n - 1$ independent W -fields.

Let $\phi^i, i = 1, \dots, n - 1$ be our free fields with correlation functions

$$\langle \phi^i(z) \phi^j(z') \rangle = -\delta^{ij} \ln(z - z') . \quad (2.1)$$

They have a mode expansion

$$\phi^j(z) = \phi_0^j - i a_0^j \ln z + i \sum_{n \neq 0} a_n^j \frac{z^{-n}}{n} \quad (2.2)$$

(where $i = \sqrt{-1}$) and the modes have the commutators

$$[a_n^j, a_m^l] = \delta^{jl} n \delta_{n+m, 0} , \quad [\phi_0^j, a_0^l] = i \delta^{jl} . \quad (2.3)$$

Thus, as usual, the zero-modes ϕ_0 and a_0 behave as conjugate position and momentum operators.

The Hilbert space is a direct sum of Fock spaces, where each Fock space has a vacuum $|\alpha\rangle$ labelled by the eigenvalue of the zero-mode a_0^j :

$$a_0^j|\alpha\rangle = \alpha^j|\alpha\rangle. \quad (2.4)$$

The states of the Fock space built on $|\alpha\rangle$ are of the form $a_{-n_1}^{j_1} \dots a_{-n_r}^{j_r}|\alpha\rangle$. In fact, below, we will impose hermiticity conditions like $a_n^+ = -a_{-n}$ ($n \neq 0$) such that the inner product is not positive definite. (Thus, strictly speaking, we are not really dealing with a Hilbert space.) Different Fock spaces are connected by the operators $e^{i\beta \cdot \phi_0}$:

$$e^{i\beta \cdot \phi_0}|\alpha\rangle = |\alpha + \beta\rangle \quad (2.5)$$

as is easily seen using the second equation (2.3).

The stress tensor

The most general form of the (holomorphic part of the) stress tensor one may construct out of $n-1$ free fields is

$$T(z) = -\frac{1}{2}\partial\phi(z) \cdot \partial\phi(z) + 2i\alpha_0\rho \cdot \partial^2\phi(z). \quad (2.6)$$

In general, the quadratic term is $-\frac{1}{2}\partial\phi^i M_{ij} \partial\phi^j$ where M is the inverse of the matrix giving the correlators between the ϕ 's in order to obtain the correct OPE. Since by eq. (2.1) $(M^{-1})_{ij} = \delta_{ij}$ the quadratic term must be as in (2.6). One is always free to add a linear term where ρ is some fixed vector with $n-1$ components. For later convenience we have pulled out a constant $2i\alpha_0$. It is an easy exercise to verify that this stress tensor (2.6) indeed satisfies the correct OPE (cf (1.5)) with central charge given by

$$c = (n-1) - 48\alpha_0^2\rho^2. \quad (2.7)$$

(Chiral) primary fields = vertex operators

Chiral vertex operators are defined by

$$V_\alpha(z) = : e^{i\alpha \cdot \phi(z)} : = e^{i\alpha \cdot \phi_0} z^{\alpha \cdot a_0} \exp\left(\alpha \cdot \sum_{n=1}^{\infty} a_{-n} \frac{z^n}{n}\right) \exp\left(-\alpha \cdot \sum_{n=1}^{\infty} a_n \frac{z^{-n}}{n}\right). \quad (2.8)$$

Due to the presence of $e^{i\alpha \cdot \phi_0}$ these vertex operators shift the a_0 -eigenvalues by α , thus interpolating between different Fock spaces. As will become clear below,

different representations of the W_n -algebra correspond to different Fock spaces (up to certain identifications), hence, in general, these vertex operators interpolate between different representations. It is a straightforward computation (using (2.1)) to show that the OPE with the stress tensor reads

$$T(z)V_\alpha(z') = \frac{\Delta_2(\alpha)V_\alpha(z')}{(z-z')^2} + \frac{\partial V_\alpha(z')}{z-z'} \quad (2.9)$$

with

$$\Delta_2(\alpha) = \frac{1}{2}\alpha \cdot (\alpha - 4\alpha_0\rho) = \frac{1}{2}(\alpha - 2\alpha_0\rho)^2 - 2\alpha_0^2\rho^2 \quad (2.10)$$

telling us that V_α is a Virasoro primary field with conformal dimension $\Delta_2(\alpha)$.

Some conventions: roots and weights of $A_{n-1} \sim su(n)$

It is convenient to introduce some specific basis in the $(n-1)$ -dimensional “target-space” of our free-fields. For this we choose the simple roots or the fundamental weights of the Lie algebra A_{n-1} (see e.g. ref. 18). Recall that all roots may be divided into positive and negative roots, and that one may choose a basic set of $n-1$ simple roots e_i such that every positive root can be written as $\sum m_i e_i$ with *all* m_i positive, and similarly for a negative root with *all* m_i negative. These simple roots may be written as $e_i = (0, \dots, 0, -1, 1, 0, \dots, 0)$ in a n -dimensional space. In this n -dimensional space the root and weight spaces lie in the $n-1$ -dimensional hyperplane defined by the vanishing of the sum of all n components. This might seem rather complicated at first sight, so we invite the reader to look at the example of the weight/root diagram of $su(3)$ shown below.

We will always normalize the roots to have length squared equal to 2, thus in particular

$$e_i^2 = 2. \quad (2.11)$$

One also introduces the $n-1$ fundamental weights λ_i dual to the simple roots:

$$\lambda_i \cdot e_j = \delta_{ij}. \quad (2.12)$$

These are the highest weights of the completely antisymmetric tensor (with i indices) representations corresponding to vertical Young tableaux with i boxes.

The Weyl vector ρ is defined as half the sum of the positive roots and equals

$$\rho = \sum_{i=1}^{n-1} \lambda_i . \quad (2.13)$$

It follows that $\rho \cdot e_i = 1$ for all i . The square of the Weyl vector can be easily computed by using the fact that $\lambda_i \cdot \lambda_j$ is given by the (i, j) element of the inverse of the Cartan matrix of A_{n-1} (i.e. $\lambda_i \cdot \lambda_j = \frac{i(n-j)}{n} \delta_{i \leq j} + \frac{j(n-i)}{n} \delta_{i < j}$) and one obtains

$$\rho^2 = \frac{(n-1)n(n+1)}{12} . \quad (2.14)$$

We also note for later use that it follows that $\rho \cdot \lambda_i = \sum_j \lambda_j \cdot \lambda_i = \frac{i(n-i)}{2} = \frac{n}{2} \lambda_i^2$.

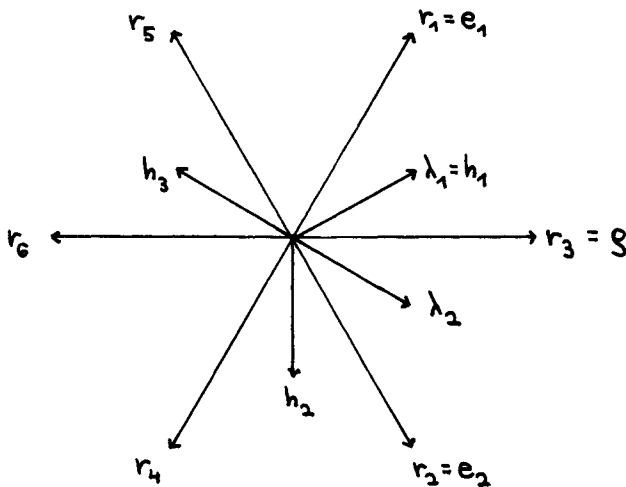


Figure: The weight/root diagram of $su(3)$.

The six non-zero roots of $su(3)$ are denoted $r_1, \dots, r_6$. We may choose r_1, r_2, r_3 as positive roots. Then $r_1 = e_1$ and $r_2 = e_2$ are the simple roots, while r_3 is the highest root. We have also drawn the two fundamental weights λ_1, λ_2 and the weights h_1, h_2, h_3 of the 3-representation having λ_1 as highest weight.

We are free to choose the vector ρ which appeared in the stress tensor to coincide with the Weyl vector. The Virasoro central charge then reads

$$c = (n-1) (1 - 4\alpha_0^2 n(n+1)) . \quad (2.15)$$

Finally, we will need the n weights of the vector representation, i.e. the one with λ_1 as highest weight. They will be called h_μ and are given explicitly by

$$h_1 = \lambda_1 , \quad h_2 = \lambda_1 - e_1 , \quad \dots , \quad h_\mu = \lambda_1 - \sum_{i=1}^{\mu-1} e_i . \quad (2.16)$$

One has the following relations: $\sum_{\mu=1}^n h_\mu = 0$, $h_\mu \cdot h_\nu = \delta_{\mu\nu} - \frac{1}{n}$, $e_i = h_i - h_{i+1}$, $\lambda_i = \sum_{\mu=1}^i h_\mu$ and $h_\mu = \lambda_\mu - \lambda_{\mu-1}$ ($\lambda_0 = \lambda_{n+1} = 0$). It then follows that the Weyl vector equals $\rho = \sum_{\mu} (n - \mu) h_\mu = - \sum_{\mu} \mu h_\mu$.

Operators of conformal dimension 1 (screening operators)

A certain set of operators with conformal dimensions equal to 1 will prove particularly important in the following. Define

$$V_{\pm}^j(z) \equiv V_{\alpha_{\pm} e_j}(z) = : e^{i\alpha_{\pm} e_j \cdot \phi(z)} : . \quad (2.17)$$

Their conformal dimensions are

$$\Delta_2(\alpha_{\pm} e_j) = \frac{1}{2}(\alpha_{\pm} e_j)^2 - 2\alpha_0 \alpha_{\pm} \rho \cdot e_j = \alpha_{\pm}^2 - 2\alpha_0 \alpha_{\pm} . \quad (2.18)$$

Thus they have unit conformal dimension provided

$$\alpha_{\pm}^2 - 2\alpha_0 \alpha_{\pm} - 1 = 0 . \quad (2.19)$$

Since this is a quadratic equation we have two solutions as already anticipated by the notation $\alpha_{\pm}$. This provides us with $2(n-1)$ operators of conformal dimension

1. Such operators are particularly interesting since one has

$$T(z)V_{\pm}^j(z') = \frac{V_{\pm}^j(z')}{(z-z')^2} + \frac{\partial V_{\pm}^j(z')}{z-z'} = \partial_{z'} \left(\frac{V_{\pm}^j(z')}{z-z'} \right) \quad (2.20)$$

which is a total derivative. Thus $\oint dz' T(z)V_{\pm}^j(z') = 0$, or equivalently

$$[L_n, Q_{\pm}^j] = 0, \quad Q_{\pm}^j = \oint dz' V_{\pm}^j(z') \quad (2.21)$$

where the integral runs over some appropriately chosen closed contour. Of course there are many more operators of unit conformal dimension, since the equation $\Delta_2(\alpha) = 1$ has an $(n-2)$ -dimensional solution-manifold. However, we are only interested in the operators (2.17) since they also commute with all $W_n^{(k)}$ as we will see below.

Construction of the higher-spin fields $W^{(k)}(z)$

Our goal is to find primary fields of conformal dimensions (spins) $k = 3, \dots, n$, built out of the free fields and forming a closed algebra. By closed we mean closed in the usual non-linear sense, i.e. the OPE of any two $W^{(k)}$, $W^{(l)}$ should be re-expressible in terms of differential polynomials of the other $W^{(r)}$'s with *no* explicit dependence on the ϕ 's. One may try to find these higher-spin fields basically by trial and error. Clearly, a good candidate for $W^{(k)}$ is some combination of $(\partial\phi)^k$, $(\partial\phi)^{k-2}\partial^2\phi$, etc, anything containing a total of k derivatives. One might write down the most general such expression and then compute the OPE with the stress tensor, determine the coefficients so as to obtain a primary field of dimension k , go on and compute the OPEs among the $W^{(k)}$'s, and so on. This clearly is reasonable only for the W_3 -algebra, or maybe for some small n using algebraic manipulation programs. However, having figured out what is $W^{(3)}$ for the W_3 -algebra, one observes an underlying structure related to the weights of $su(3)$. This motivates the following guess: the $W^{(k)}$'s are obtained by considering the following differential operator of order n .

$$\begin{aligned} (2i\alpha_0)^n \mathcal{D}_n &= : \prod_{\mu=1}^n (2i\alpha_0 \partial_z + h_{\mu} \cdot \partial \phi(z)) : \\ &\equiv: (2i\alpha_0 \partial_z + h_n \cdot \partial \phi(z)) \dots (2i\alpha_0 \partial_z + h_2 \cdot \partial \phi(z)) (2i\alpha_0 \partial_z + h_1 \cdot \partial \phi(z)) : . \end{aligned} \quad (2.22)$$

This operator naturally appears, e.g. in the study of Toda theories. Now expand

this differential operator as

$$\mathcal{D}_n = \partial_z^n + \sum_{k=1}^n (2i\alpha_0)^{-k} u_k(z) \partial_z^{n-k} . \quad (2.23)$$

One finds:

$$u_1 = \sum_{\mu} h_{\mu} \cdot \partial \phi = 0 \quad (2.24)$$

by the properties of the weights of the vector representation (this is essentially due to the fact that the trace in any representation vanishes since we are dealing with $su(n)$). Next

$$u_2 = \sum_{\mu > \nu} : h_{\mu} \cdot \partial \phi h_{\nu} \cdot \partial \phi : + 2i\alpha_0 \sum_{\mu} (n - \mu) h_{\mu} \cdot \partial^2 \phi . \quad (2.25)$$

Rewriting the first sum as

$$\sum_{\mu > \nu} = \frac{1}{2} \sum_{\mu \neq \nu} = \frac{1}{2} \sum_{\mu, \nu} - \frac{1}{2} \sum_{\mu = \nu} = \frac{1}{2} \left(\sum_{\mu} \right)^2 - \frac{1}{2} \sum_{\mu = \nu} = 0 - \frac{1}{2} \sum_{\mu = \nu} \quad (2.26)$$

and recognizing $\sum_{\mu} (n - \mu) h_{\mu} = \rho$, one finally arrives at

$$u_2 = -\frac{1}{2} : \partial \phi(z) \cdot \partial \phi(z) : + 2i\alpha_0 \rho \cdot \partial^2 \phi(z) = T \quad (2.27)$$

which coincides with the stress tensor. This is encouraging, and one might expect that the other u_k will yield the $W^{(k)}$. This is almost true. In fact one has e.g. (we do not write the normal ordering explicitly any longer)

$$\begin{aligned} u_3 = & \sum_{\mu > \nu > \sigma} h_{\mu} \cdot \partial \phi h_{\nu} \cdot \partial \phi h_{\sigma} \cdot \partial \phi + 2i\alpha_0 \sum_{\mu > \nu} h_{\mu} \cdot \partial \phi (n - 1 - \nu) h_{\nu} \cdot \partial^2 \phi \\ & + 2i\alpha_0 \sum_{\nu > \mu} (n - \nu) h_{\nu} \cdot \partial^2 \phi h_{\mu} \cdot \partial \phi + (2i\alpha_0)^2 \sum_{\mu} \binom{n - \mu}{2} h_{\mu} \cdot \partial^3 \phi . \end{aligned} \quad (2.28)$$

One then computes the OPE of T with this u_3 and one finds that u_3 almost behaves like a spin 3 primary field. It is a quasi-primary field (i.e. the OPE contains u_3

and ∂u_2 on its r.h.s.). This however can be cured by defining

$$W^{(3)} = u_3 - \frac{n-2}{2}(2i\alpha_0)\partial u_2. \quad (2.29)$$

This $W^{(3)}$ is a true primary spin 3 field. Computing its OPE with itself yields exactly, for $n = 3$, the OPE of the W_3 -algebra I have written down in the introduction.* Similarly one finds for general n that u_4 is only quasiprimary, but again a true primary spin 4 field can be defined as

$$W^{(4)} = u_4 + \beta\partial u_3 + \gamma\partial^2 u_2 + \delta : u_2^2 : \quad (2.30)$$

with some numerical coefficients β, γ, δ .

Two questions arise:

1. Is it always possible to define true primary fields $W^{(k)}$ out of the u_k for all $k \leq n$?
2. Does the OPE algebra always close in the usual non-linear sense (on differential polynomials of the $W^{(r)}$'s)? It is clearly enough to answer this for the u_k 's.

In principle, for any given n the answer will be obtained “just” by computing the OPEs. However one would like to have some answer for generic n . In the *classical* case, i.e. disregarding normal ordering effects (multiple contractions between the free fields) which is equivalent to computing Poisson brackets, the answer to both questions is “yes” [8,19]. In the present quantum case these questions are much more difficult to answer. For an affirmative answer to the second question see ref. 20 or ref. 17. For the rest of these lectures it will not matter whether or not one can define true primary fields $W^{(k)}$ out of the u_k also in the quantum theory. Thus $W^{(k)}$ will merely stand for *some* (well-defined) combination $W^{(k)} = u_k + \beta'\partial u_{k-1} + \gamma'\partial u_{k-2} + \delta' : u_2 u_{k-2} + \dots$, be it primary or quasiprimary.

* To be precise, the $W^{(3)}$ in the introduction was normalized differently: $(W^{(3)})_{\text{introduction}} = (-2/5)^{1/2}(2i\alpha_0)^{-1}(W^{(3)})_{\text{eq.(2.29)}}$ (for $n = 3$).

2.2 HIGHEST WEIGHT REPRESENTATIONS OF THE W_n -ALGEBRAS

The relation between the highest weights and the α 's

Recall that the vertex operator $V_\alpha(z) = : e^{i\alpha \cdot \phi(z)} :$ is a Virasoro primary field with conformal dimension $\Delta_2(\alpha) = \frac{1}{2}\alpha \cdot (\alpha - 4\alpha_0\rho)$. Let $|0\rangle$ be the Fock space vacuum with a_0 -eigenvalue 0. Thus $a_n^j|0\rangle = 0$, $\forall n \geq 0$. It is easily seen that one also has $L_n|0\rangle = 0$, $\forall n \geq -1$, so this is the $Sl(2, C)$ -invariant vacuum state. Since any positive or zero mode of $W^{(k)}$ always contains some oscillator a_m^j with positive or zero m , it is clear that is also annihilated by the $W_m^{(k)}$, $\forall m \geq 0$. As usual the highest weight state with highest weight $\Delta_2(\alpha)$ is obtained by applying V_α to the $Sl(2, C)$ -invariant vacuum:

$$\lim_{z \rightarrow 0} V_\alpha(z)|0\rangle = e^{i\alpha \cdot \phi_0}|0\rangle = |\alpha\rangle \quad (2.31)$$

It is again clear that this state is annihilated by all positive modes of the $W^{(k)}$ and of T , and thus is a highest weight state of the W_n -algebra:

$$\begin{aligned} W_n^{(k)}|\alpha\rangle &= 0, \quad \forall n > 0 \\ W_0^{(k)}|\alpha\rangle &= \bar{\Delta}_k(\alpha)|\alpha\rangle, \quad k = 2, 3, \dots, n \end{aligned} \quad (2.32)$$

(where $W^{(2)}$ denotes T). It is an easy exercise to compute the values of $\bar{\Delta}_k(\alpha)$. In fact, we will compute the eigenvalues of the zero modes of the $u_k(z) = \sum_n u_n^k z^{-n-k}$ which we will call $\Delta_k(\alpha)$. The $\bar{\Delta}_k(\alpha)$ and $\Delta_k(\alpha)$ are related by eqs of the type (2.29), (2.30), e.g. $\bar{\Delta}_3(\alpha) = \Delta_3(\alpha) + (n-2)(2i\alpha_0)\Delta_2(\alpha)$ and, of course, $\bar{\Delta}_2(\alpha) = \Delta_2(\alpha)$. Consider the differential operator (with operator-valued coefficients) $\mathcal{D}_n$ defined above and apply the highest weight state $|\alpha\rangle$ on the right:

$$\begin{aligned} (2i\alpha_0)^n \left(\partial^n + \sum_{k=2}^n \frac{\Delta_k(\alpha)}{(2i\alpha_0)^k z^k} \partial^{n-k} \right) |\alpha\rangle \\ = \left(2i\alpha_0 \partial - i \frac{h_n \cdot a_0}{z} \right) \dots \left(2i\alpha_0 \partial - i \frac{h_1 \cdot a_0}{z} \right) |\alpha\rangle \end{aligned} \quad (2.33)$$

where we only wrote the most singular terms as $z \rightarrow 0$. The left hand and right hand sides are still differential operators which we apply on the monomials

z^j , $j = 0, \dots, n-2$ to obtain $n-2$ coupled linear algebraic equations for the $\Delta_k(\alpha)$:

$$\sum_{k=0}^j (2i\alpha_0)^k \frac{j!}{(j-k)!} \Delta_{n-k}(\alpha) = i^n \prod_{m=1}^n (2\alpha_0(j-m+1) - h_m \cdot \alpha) \quad , \quad j = 0, \dots, n-2 \quad (2.34)$$

with solution

$$\Delta_k(\alpha) = (-i)^k \sum_{\mu_1 > \mu_2 > \dots > \mu_k} \prod_{m=1}^k (h_{\mu_m} \cdot \alpha + 2\alpha_0(k-m)) \quad (2.35)$$

and one easily checks that this correctly reproduces $\Delta_2(\alpha) = \frac{1}{2}\alpha^2 - 2\alpha_0\rho \cdot \alpha$.

An important feature one sees from eq. (2.35) is that the correspondence between α and the set $(\Delta_2, \Delta_3, \dots, \Delta_n)$ is *not* one-to-one, but many-to-one. This is already clear for $n = 2$ where α and $4\alpha_0\rho - \alpha$ lead to the same Δ_2 . From a representation-theoretic point of view this means that all V_α with the same Δ_k 's have to be identified. Now, the r.h.s of eq. (2.34) is unchanged if the effect of replacing α by some $\tilde{\alpha}$ is simply a permutation of the factors of the product, i.e. if

$$h_m \cdot \alpha + 2\alpha_0 m = h_{\pi(m)} \cdot \tilde{\alpha} + 2\alpha_0 \pi(m) \quad (2.36)$$

where π is a permutation of $(1, 2, \dots, n)$. Using $h_m \cdot \rho = \frac{n+1}{2} - m$ one can rewrite this condition as

$$h_m \cdot (\alpha - 2\alpha_0\rho) = h_{\pi(m)} \cdot (\tilde{\alpha} - 2\alpha_0\rho) \quad (2.37)$$

Clearly, there are $n!$ possible permutations of $(1, 2, \dots, n)$, and the relation between α and the set $(\Delta_2, \Delta_3, \dots, \Delta_n)$ is $n!$ -to-one. For later use it is convenient to consider the special permutation $(1, 2, \dots, n) \rightarrow (n, \dots, 2, 1)$. For any α define α^* by

$$h_m \cdot \alpha^* = -h_{n+1-m} \cdot \alpha \quad , \quad m = 1, \dots, n \quad (2.38)$$

Then one has $\rho^* = \rho$ and $h_m \cdot ((4\alpha_0\rho - \alpha^*) - 2\alpha_0\rho) = h_{n+1-m} \cdot (\alpha - 2\alpha_0\rho)$, and thus

$$\Delta_k(4\alpha_0\rho - \alpha^*) = \Delta_k(\alpha) \quad , \quad \forall k \quad (2.39)$$

Note also for further use that $e_j \cdot \alpha^* = (h_j - h_{j+1}) \cdot \alpha^* = e_{n-j} \cdot \alpha$ and $e_j^* = e_{n-j}$, $\lambda_j^* = \lambda_{n-j}$. Since the star-operation only permutes the e_j it preserves the length

of any vector^{*} and hence $(\alpha^* - 2\alpha_0\rho)^2 = (\alpha^* - 2\alpha_0\rho^*)^2 = (\alpha - 2\alpha_0\rho)^2$. It follows that $\Delta_2(\alpha^*) = \Delta_2(\alpha)$.

At this point we would like to comment on the hermiticity properties in the Hilbert space. Since $L_n = a_n \cdot a_0 + \dots - 2\alpha_0(n+1)\rho \cdot a_n$ one has $L_n^+ = L_{-n}$ provided

$$a_n^+ = -a_{-n} \quad \text{and} \quad a_0^+ = 4\alpha_0\rho - a_0. \quad (2.40)$$

In fact, this also leads to $W_n^{(k)+} = W_{-n}^{(k)}$. The relation $a_0^+ = 4\alpha_0\rho - a_0$ implies that the state conjugate to $|\alpha\rangle$ is $\langle 4\alpha_0\rho - a_0|$:

$$|\alpha\rangle^+ = \langle 4\alpha_0\rho - a_0|. \quad (2.41)$$

It is clear that the hermiticity properties (2.40) imply that the scalar product is indefinite and there are many states with negative norm, like e.g. $a_{-1}^j|\alpha\rangle$. Thus, in general, the representations of the W -algebras we are considering are not unitary. It is only for certain values of the central charge that they will turn out to be unitary (see below).

It is useful to get a better understanding of the transformation $\alpha \rightarrow \tilde{\alpha}$ defined by eq. (2.37) for a general permutation π . It is easy to see that any permutation of the weights h_μ corresponds to an element of the Weyl group (the group generated by the reflections by the hyperplanes normal to the roots). An arbitrary weight w can be written as $w = \sum_i l_i \lambda_i$. If all $l_i \geq 0$ the weight w is called a dominant (or highest) weight. Any weight w can be mapped into a (unique) dominant weight λ_w by an element of the Weyl group. Note that the Weyl vector is a dominant weight. As we will show below, the “interesting” (i.e. completely degenerate) representations of the W -algebra have $\alpha = -\alpha_- \lambda - \alpha_+ \lambda'$ with λ and λ' dominant weights. Since $2\alpha_0 = \alpha_- + \alpha_+$, one has $\alpha - 2\alpha_0\rho = -\alpha_-(\lambda + \rho) - \alpha_+(\lambda' + \rho)$. Equation (2.37) tells us that there are $n!$ equivalent representatives $\tilde{\alpha}$ of α where $\tilde{\lambda} + \rho$ is obtained from $\lambda + \rho$ by applying an element of the Weyl group, and $\tilde{\lambda}' + \rho$ from $\lambda' + \rho$ by applying the same element. For later use, we also remark the following: the generators of Weyl transformations are written explicitly as

* Indeed, if $v = \sum_{i=1}^{n-1} v_i e_i$ then clearly $v^2 = 2 \sum_{i=1}^{n-1} v_i^2 - \sum_{i=1}^{n-2} v_i v_{i+1} - \sum_{i=2}^{n-1} v_i v_{i-1} = v^{*2}$ where $v^* = \sum_{i=1}^{n-1} v_i e_{n-i}$.

$w \rightarrow \hat{w} = w - 2\frac{w \cdot e}{e^2}e = w - w \cdot e e$ where e is a root. Since $w \cdot e$ is an integer, $w - \hat{w}$ is always a root. If w is a dominant weight, $\hat{w} = w - r_w$ where r_w is a *positive* root. Since the Weyl vector is a dominant weight, we also have $\hat{\rho} = \rho - r_\rho$, r_ρ a positive root. It then follows that $\tilde{\lambda} + \rho = \hat{\lambda} + \hat{\rho} = \lambda - r_\lambda + \rho - r_\rho$. Thus $\tilde{\lambda} = \lambda - r$, r a positive root.

This is the generic situation. Of course, if α_+/α_- is not generic (i.e. irrational) but rational, $\alpha_+/\alpha_- = -\alpha_+^2 = -p'/p$, one has $\alpha - 2\alpha_0\rho = (p(\lambda + \rho) - p'(\lambda' + \rho))/\sqrt{pp'}$. Thus, effectively, α is not labelled by the weights λ and λ' separately but only by the combination $\Lambda = p(\lambda + \rho) - p'(\lambda' + \rho)$. As a consequence, there are many more transformations $(\lambda, \lambda') \rightarrow (\tilde{\lambda}, \tilde{\lambda}')$ that result in a Weyl transformation of the weight Λ , and further field identifications (of the V_α) have to be performed. This leads to the minimal models to be discussed in more detail below.

The role of the screening operators

Recall the form of the stress tensor, $T = -\frac{1}{2}\partial\phi \cdot \partial\phi + 2i\alpha_0\rho \cdot \partial^2\phi$. The quadratic term is the standard piece for $n - 1$ free fields. The term linear in $\partial^2\phi$ can be interpreted as being due to a “background” charge. Indeed, the stress tensor is the variation of the action with respect to the metric. It is easy to see that the linear term is obtained if one adds to the standard free field action a term

$$S_0 \rightarrow S_0 + \frac{i\alpha_0}{4\pi} \int \sqrt{g} R \rho \cdot \phi. \quad (2.42)$$

Consider the computation of a correlation function in the functional integral approach

$$\langle \prod_i V_{\alpha_i} \rangle \sim \int \mathcal{D}\phi \left(\prod_i V_{\alpha_i} \right) \exp \left(-S_0 - \frac{i\alpha_0}{4\pi} \int \sqrt{g} R \rho \cdot \phi \right). \quad (2.43)$$

These correlation functions obey a very simple Ward identity derived by changing variables $\phi^j \rightarrow \phi^j + c^j$. Thus (since $S_0 \sim \int \partial\phi \bar{\partial}\phi$ is unchanged)

$$\langle \prod_i V_{\alpha_i} \rangle = \langle \prod_i V_{\alpha_i} \rangle \exp \left(i \sum_j \alpha_j \cdot c - \frac{i\alpha_0}{4\pi} \int \sqrt{g} R \rho \cdot c \right). \quad (2.44)$$

This must be true for arbitrary c^j , and thus the correlation function can be non-

vanishing only if

$$\sum_j \alpha_j = \frac{\alpha_0}{4\pi} \int \sqrt{g} R \rho = 2\chi \alpha_0 \rho \quad (2.45)$$

by the Gauss-Bonnet theorem. χ is the Euler characteristic of the Riemann surface and equals $2(1-g)$, g being the number of handles (the genus). Thus on the sphere, correlation functions are non-vanishing only if

$$\sum_j \alpha_j = 4\alpha_0 \rho \quad (\text{on the sphere}) \quad (2.46)$$

The presence of the extra term in the action also leads to a non-conservation of the current $\partial_\mu \phi$ by an amount proportional to α_0 , hence the name "background" charge.

Non-zero two point functions on the sphere are given by $\langle V_{4\alpha_0\rho-\alpha} V_\alpha \rangle \sim \langle 4\alpha_0\rho - \alpha | \alpha \rangle \neq 0$ by eq. (2.41). Note that these are precisely the two-point functions that satisfy the charge conservation condition (2.46). According to the above discussion $V_{4\alpha_0\rho-\alpha}$ should be identified with V_{α^*} . Hence the non-zero two-point functions are

$$\langle V_{\Delta(\alpha^*)} V_{\Delta(\alpha)} \rangle \neq 0. \quad (2.47)$$

Although in general $\Delta_k(\alpha^*) \neq \Delta_k(\alpha)$ for $k \geq 3$, we noted above that $\Delta_2(\alpha^*) = \Delta_2(\alpha)$. Thus the two operators involved in the non-zero two-point function have the same conformal dimension as required by general arguments [21].

The appearance of α^* instead of α can be understood by considering the following example. Let $\alpha \sim \lambda_1$. Then, roughly speaking (this will be made more precise below) V_α corresponds to the representation with highest weight λ_1 (the 3-representation for $su(3)$). In order that $\langle V_{\Delta(\beta)} V_{\Delta(\alpha)} \rangle$ be non vanishing, the product of the representations corresponding to α and β must contain the trivial representation. This is precisely the case if $\beta \sim \lambda_n = \lambda_1^*$ (corresponding to the $\bar{3}$ -representation for $su(3)$). Thus one should have $\beta = \alpha^*$.

Consider now higher-point correlation functions. In general the charge balance condition (2.46) will not be fulfilled, although one expects the correlation function to be non-vanishing. However, one may always insert some operators that do not change the conformal (and more generally the W -algebraic) properties,

but nevertheless carry some non-zero amount of charge. These operators having zero conformal dimensions cannot be local (otherwise they would be the identity operator), but one can write down integrals of dimension-one operators that have the required properties. These are precisely the screening operators we already discussed above. We will show below that they commute with all W -generators. Inserting m_i screening operators V_-^i and m'_i operators V_+^i modifies the charge balance equation as

$$\sum_j \alpha_j + \sum_{i=1}^{n-1} (m_i \alpha_- + m'_i \alpha_+) e_i = 4\alpha_0 \rho \quad (2.48)$$

For $n = 2$ (Virasoro) one may obtain the “allowed” values of the α ’s by considering the four-point functions $\langle V_\alpha V_\alpha V_\alpha V_{4\alpha_0\rho-\alpha} \rangle$ [22]. However, for general n the argument is less convincing. We will obtain the spectrum of α -values below by studying the existence of null vectors using the screening operators.

Degenerate representations - null vectors

The Verma module $\mathcal{V}_\Delta$ is defined as the module spanned by all vectors of the type $\prod_{k=2}^n \prod_r W_{-n_r}^{(k)} |\Delta \rangle$ ($n_r^k > 0$) where $|\Delta \rangle$ denotes the highest weight state of the representation satisfying

$$W_n^{(k)} |\Delta \rangle = 0 \quad \forall n > 0, \quad W_0^{(k)} |\Delta \rangle = \bar{\Delta}_k |\Delta \rangle \quad (2.49)$$

(As discussed above, in the free-field Hilbert space this highest weight state has $n!$ representatives $|\alpha \rangle$ with $\Delta_k(\alpha) = \Delta_k$, but at present we only discuss the representation theoretic point of view, and there is only one highest weight state for a given set of Δ_k ’s.) The Verma module $\mathcal{V}_\Delta$ contains a null vector $|\chi_N \rangle$ at level N if $|\chi_N \rangle$ itself is a highest weight vector and $L_0 |\chi_N \rangle = (\Delta_2 + N) |\chi_N \rangle$. Thus $|\chi_N \rangle$ is at the same time a descendent and a highest weight state, meaning that the representation carried by the Verma module cannot be irreducible. As usual one shows that a null vector (and its descendents) decouple from the rest of the Verma module. Let e.g. $|v \rangle \in \mathcal{V}_\Delta$ i.e. $|v \rangle = \dots W_{-m_i}^{(k)} \dots |\Delta \rangle$; then $\langle v | \chi_N \rangle = \langle \Delta | \dots W_{m_i}^{(k)} \dots | \chi_N \rangle = 0$. Thus one can consistently set $|\chi_N \rangle = 0$. As an example consider the W_3 -algebra (with the generators normalized as in the

introduction). There is a null vector at level 1

$$|\chi_1\rangle = \left(2\Delta_2 W_{-1}^{(3)} - 3\bar{\Delta}_3 L_{-1}\right) |\Delta\rangle \quad (2.50)$$

provided Δ_2 and $\bar{\Delta}_3$ are related by

$$9\bar{\Delta}_3^2 = 2\Delta_2^2 \left(\frac{32}{22+5c} \left(\Delta_2 + \frac{1}{5} \right) - \frac{1}{5} \right). \quad (2.51)$$

If the Verma module $\mathcal{V}_\Delta$ contains a null vector the representation with highest weight Δ is said to be degenerate. If it contains $2(n-1)$ null vectors it is said to be completely degenerate.

The existence of null vectors implies conditions for correlation functions. For the null vector just discussed one has e.g.

$$0 = \langle \Phi_1 \dots \Phi_N \left(2\Delta_2 W_{-1}^{(3)} - 3\bar{\Delta}_3 L_{-1} \right) |\Delta\rangle \quad (2.52)$$

since we have set $|\chi_1\rangle$ to zero. Provided one knows how to commute the $W_{-1}^{(3)}$ with the Φ_i until one reaches the vacuum on the left, this will yield differential equations for the correlation function $\langle \Phi_1 \dots \Phi_N \Phi_\Delta \rangle$. Theories where all Verma modules contain null vectors are particularly interesting since then all correlation functions obey differential equations. Completely degenerate representations contain enough null vectors to completely determine the correlation functions.

Screening operators lead to null vectors

Following ref. 16 we are going to prove the following

Lemma: Recall that the screening operators were defined as $V_\pm^j =: e^{i\alpha_\pm e_j \cdot \phi(z)} :$ with $\alpha_\pm$ satisfying $\alpha_\pm^2 - 2\alpha_0\alpha_\pm - 1 = 0$. Then

$$W^{(k)}(z) V_\pm^j(z') = \mathcal{I}_{z'}(\dots) + \mathcal{O}(1) \quad (2.53)$$

(where $\mathcal{O}(1)$ denotes terms regular as $z \rightarrow z'$) or equivalently if one defines $Q_\pm^j = \oint dz' V_\pm^j(z')$ for some closed contour enclosing z

$$[W^{(k)}(z), Q_\pm^j] = 0 \quad (2.54)$$

Proof: Of course, it is enough to show the corresponding statement for the $u_k(z)$. This in turn is equivalent to showing that for the differential operator $\mathcal{D}_n$ having

the $u_k(z)$ as coefficients one has $\mathcal{D}_n(z)V_{\pm}^j(z') = \partial_{z'}(\dots) + \mathcal{O}(1)$. But recall that

$$(2i\alpha_0)^n \mathcal{D}_n = : \dots (2i\alpha_0 \partial_z + h_{j+1} \cdot \partial \phi(z)) (2i\alpha_0 \partial_z + h_j \cdot \partial \phi(z)) \dots : \quad (2.55)$$

Since $h_\mu = \lambda_\mu - \lambda_{\mu-1}$ one has $e_j \cdot h_\mu = \delta_{j,\mu} - \delta_{j+1,\mu}$. Thus the only non-zero contractions of $e_j \cdot \phi(z')$ in $V_{\pm}^j(z')$ are with the two factors in $\mathcal{D}_n$ we have written out explicitly. Using $\langle h_\mu \cdot \partial \phi(z) e_j \cdot \phi(z') \rangle = (\delta_{\mu,j} - \delta_{\mu,j+1}) \frac{-1}{z-z'}$ one finds

$$\begin{aligned} & : (2i\alpha_0 \partial_z + h_{j+1} \cdot \partial \phi(z)) (2i\alpha_0 \partial_z + h_j \cdot \partial \phi(z)) : V_{\pm}^j(z') \\ &= : \left(2i\alpha_0 \partial_z \frac{-i\alpha_{\pm}}{z-z'} + h_{j+1} \cdot \partial \phi(z) \frac{-i\alpha_{\pm}}{z-z'} + h_j \cdot \partial \phi(z) \frac{i\alpha_{\pm}}{z-z'} + \frac{\alpha_{\pm}^2}{(z-z')^2} \right) V_{\pm}^j(z') : \\ & \quad + \text{regular as } z \rightarrow z' \\ &= : \left(\frac{-2\alpha_0 \alpha_{\pm} + \alpha_{\pm}^2}{(z-z')^2} + \frac{\partial_{z'}}{z-z'} \right) V_{\pm}^j(z') : + \text{regular as } z \rightarrow z' \\ &= \partial_{z'} \left(\frac{V_{\pm}^j(z')}{z-z'} \right) + \text{regular as } z \rightarrow z' \end{aligned} \quad (2.56)$$

where we first used again $h_j - h_{j+1} = e_j$ and then the quadratic equation satisfied by $\alpha_{\pm}$. This completes the proof.

Now it is an easy exercise to construct null vectors. Since $W_m^{(k)}|\Delta(\beta)\rangle = 0 \quad \forall m > 0$ it follows that

$$W_m^{(k)} \left(Q_{\pm}^j |\Delta(\beta)\rangle \right) = Q_{\pm}^j W_m^{(k)} |\Delta(\beta)\rangle = 0 \quad \forall m > 0. \quad (2.57)$$

Thus

$$|\chi_{\pm}^j(\beta)\rangle \equiv Q_{\pm}^j |\Delta(\beta)\rangle \quad (2.58)$$

is a highest weight vector. Since $|\Delta(\beta)\rangle = V_{\beta}(0)|0\rangle$,

$$\begin{aligned} |\chi_{\pm}^j(\beta)\rangle &= \oint dz : e^{i\alpha_{\pm} e_j \cdot \phi(z)} : : e^{i\beta \cdot \phi(0)} : |0\rangle \\ &= \oint dz z^{\alpha_{\pm} e_j \cdot \beta} : e^{i\alpha_{\pm} e_j \cdot \phi(z) + i\beta \cdot \phi(0)} : |0\rangle \end{aligned} \quad (2.59)$$

with the normal ordered product being regular as $z \rightarrow 0$. This expression is

well-defined and non-zero if and only if

$$\alpha_{\pm} e_j \cdot \beta = -\bar{l}_j - 1, \quad \bar{l}_j = 0, 1, 2, \dots \quad (2.60)$$

since then

$$|\chi_{\pm}^j(\beta)\rangle = \frac{1}{\bar{l}_j!} \partial_z^{\bar{l}_j} : e^{i\alpha_{\pm} e_j \cdot \phi(z) + i\beta \cdot \phi(0)} : |0\rangle \Big|_{z=0} \quad (2.61)$$

This clearly shows that $|\chi_{\pm}^j(\beta)\rangle$ is a descendent of the highest weight state $|\Delta(\beta + \alpha_{\pm} e_j)\rangle$. Since it is also highest weight itself, it is a null vector. Now set $\beta = 4\alpha_0 \rho - \alpha^* - \alpha_{\pm} e_j$ such that $|\chi_{\pm}^j(\beta)\rangle$ is a null vector for the Verma module based on the highest weight $\Delta(4\alpha_0 \rho - \alpha^*) = \Delta(\alpha)$. Inserting the value of β into eq. (2.60) and using the quadratic relation satisfied by $\alpha_{\pm}$ we find $\alpha_{\pm} e_j \cdot \alpha^* = \bar{l}_j - 1$ or, since $\alpha_- \alpha_+ = -1$ we have $e_j \cdot \alpha^* = \alpha_{\mp}(1 - \bar{l}_j)$. Since $e_j \cdot \alpha^* = e_{n-j} \cdot \alpha$ we can rewrite this as ($l_{n-j} = \bar{l}_j$)

$$e_{n-j} \cdot \alpha = \alpha_{\mp}(1 - l_{n-j}). \quad (2.62)$$

Note that if $l_{n-j} = 0$ the would-be null vector $|\chi_{\pm}^j(\beta)\rangle$ coincides with the highest weight state $|\Delta(\alpha)\rangle$ and there is no null vector. Thus we restrict ourselves to $l_{n-j} \geq 1$.

One may similarly consider $|\chi_{\pm}^{j, \bar{l}'}\rangle = \oint \dots \oint V_{\pm}^j \dots V_{\pm}^{\bar{l}'} |\Delta(\beta)\rangle$ with $\bar{l}'_j$ screening operators. The condition that this is well-defined and non-vanishing (hence a null vector) is now [16]

$$\alpha_{\pm}^2(\bar{l}'_j - 1) + \alpha_{\pm} e_j \cdot (4\alpha_0 \rho - \alpha^* - \alpha_{\pm} e_j) = -\bar{l}_j - 1. \quad (2.63)$$

This can be rewritten as ($l_{n-j} = \bar{l}_j$, $l'_{n-j} = \bar{l}'_j$)

$$e_{n-j} \cdot \alpha = \alpha_{\mp}(1 - l_{n-j}) + \alpha_{\pm}(1 - l'_{n-j}). \quad (2.64)$$

We will have a completely degenerate representation if this is true for all j . Then,

using $\lambda_i \cdot e_j = \delta_{ij}$, the general expression for α satisfying these equations is

$$\alpha = \sum_{j=1}^{n-1} ((1 - l_j)\alpha_- + (1 - l'_j)\alpha_+) \lambda_j . \quad (2.65)$$

where $l_j, l'_j = 1, 2, \dots$. Note that one also has

$$\alpha - 2\alpha_0\rho = - \sum_{j=1}^{n-1} (l_j\alpha_- + l'_j\alpha_+) \lambda_j . \quad (2.66)$$

Thus we have arrived at the conclusion that the completely degenerate representations are those with highest weights $\Delta(\alpha)$ with α given by the preceding equation. Of course we have not proven that there are no other completely degenerate representations, but various consistency checks indicate that this is the case [16].

Note that $\lambda = \sum_j (l_j - 1)\lambda_j$ and $\lambda' = \sum_j (l'_j - 1)\lambda_j$ are highest (i.e. dominant) weights of $su(n)$. Thus we may rewrite eq. (2.65) as

$$\alpha = -\alpha_- \lambda - \alpha_+ \lambda' . \quad (2.67)$$

A W_n -highest weight is labelled by a pair (λ, λ') of highest weights of $su(n)$. This will find a very natural interpretation when we consider the coset construction below.

2.3 OPERATOR PRODUCT ALGEBRA - FUSION RULES

We now wish to discuss the fusion rules for the primary fields. Let us start with a simple remark. The OPE of two vertex operators is

$$\begin{aligned} V_\alpha(z)V_\beta(z') &= (z - z')^{\alpha \cdot \beta} : V_\alpha(z)V_\beta(z') : \\ &= (z - z')^{\alpha \cdot \beta} V_{\alpha+\beta}(z') + \text{less singular terms} . \end{aligned} \quad (2.68)$$

Thus the leading term in the OPE of two vertex operators is very easy to obtain. However, we are also interested in the subleading terms, as far as they are primary fields. Another subtlety is due to the fact that several α 's may correspond to the same highest weights and that the corresponding V_α 's must be identified as discussed above.

To find the fusion rules we will consider three-point functions

$$< V_{\Delta(\alpha^*)} V_{\Delta(\beta)} V_{\Delta(\gamma)} Q^+ \dots Q^- \dots > \quad (2.69)$$

(where the Q 's are the integrated screening operators and α^* was defined in eq. (2.38)) and ask when the charge conservation condition (2.48) can be satisfied in order to yield a non-zero three-point function. First recall from (2.47) that the non-zero two-point functions are

$$< V_{\Delta(\alpha^*)} V_{\Delta(\alpha)} > . \quad (2.70)$$

Thus, whenever the three-point function (2.69) is non-zero, we conclude that the fusion of the primary field $\Phi_{\Delta(\beta)}$ with the primary field $\Phi_{\Delta(\gamma)}$ yields the primary field $\Phi_{\Delta(\alpha)}$ (*not* $\Phi_{\Delta(\alpha^*)}$). Consider the following representation of the three-point function (2.69)

$$< V_{4\alpha_0\rho-\alpha} V_{\beta} V_{\gamma} Q^+ \dots Q^- \dots > . \quad (2.71)$$

Charge conservation tells us that

$$4\alpha_0\rho - \alpha + \beta + \gamma + \sum_{i=1}^{n-1} (m_i\alpha_- + m'_i\alpha_+)e_i = 4\alpha_0\rho . \quad (2.72)$$

Now let $\alpha = -\alpha_- \lambda_{\alpha} - \alpha_+ \lambda'_{\alpha}$, $\beta = -\alpha_- \lambda_{\beta} - \alpha_+ \lambda'_{\beta}$ and $\gamma = -\alpha_- \lambda_{\gamma} - \alpha_+ \lambda'_{\gamma}$. Thus we also have $\alpha^* = -\alpha_- \lambda_{\alpha}^* - \alpha_+ \lambda'^*_{\alpha}$, etc. The charge conservation now becomes (for *generic* values of $\alpha_{\pm}$)

$$\lambda_{\alpha} = \lambda_{\beta} + \lambda_{\gamma} - r \quad \text{and} \quad \lambda'_{\alpha} = \lambda'_{\beta} + \lambda'_{\gamma} - r' \quad (2.73)$$

where $r = \sum m_i e_i$, $r' = \sum m'_i e_i$ are *positive or zero* roots. The case $r = r' = 0$ corresponds to the leading term in (2.68). Note that equations (2.73) are part of the rules for the decomposition of the tensor product of two $su(n)$ -representations with highest weights λ_{β} and λ_{γ} into representations with highest weights λ_{α} . Since the space of roots has a unique decomposition into positive and negative roots (unique once we have chosen our set of simple roots) it is meaningful to write

$\lambda - \mu \geq 0$ to indicate that $\lambda - \mu$ is a positive or zero root. Then we may restate (2.73) as

$$\lambda_\alpha \leq \lambda_\beta + \lambda_\gamma \quad \text{and} \quad \lambda'_\alpha \leq \lambda'_\beta + \lambda'_\gamma . \quad (2.74)$$

We might have represented the three-point function (2.69) in a different way:

$$< V_{\alpha^*} V_{4\alpha_0\rho-\beta^*} V_\gamma Q^+ \dots Q^- \dots > . \quad (2.75)$$

Then the analogue of condition (2.74) reads

$$\lambda_\alpha^* \geq \lambda_\beta^* - \lambda_\gamma \quad \text{and} \quad \lambda'^*_\alpha \geq \lambda'^*_\beta - \lambda'_\gamma . \quad (2.76)$$

Finally, considering

$$< V_{\alpha^*} V_\beta V_{4\alpha_0\rho-\gamma^*} Q^+ \dots Q^- \dots > \quad (2.77)$$

leads to

$$\lambda_\alpha^* \geq \lambda_\gamma^* - \lambda_\beta \quad \text{and} \quad \lambda'^*_\alpha \geq \lambda'^*_\gamma - \lambda'_\beta . \quad (2.78)$$

Since the conditions (2.74), (2.76) and (2.78) must be all true we find the fusion rules

$$\left[\Phi_{\lambda_\beta} \lambda'_\beta \right] \times \left[\Phi_{\lambda_\gamma} \lambda'_\gamma \right] = \sum_{\substack{\lambda_\alpha^* \leq \lambda_\beta + \lambda_\gamma \\ \lambda_\alpha^* \geq \lambda_\beta^* - \lambda_\gamma \\ \lambda_\alpha^* \geq \lambda_\gamma^* - \lambda_\beta}} \sum_{\substack{\lambda'_\alpha \leq \lambda'_\beta + \lambda'_\gamma \\ \lambda'^*_\alpha \geq \lambda'^*_\beta - \lambda'_\gamma \\ \lambda'^*_\alpha \geq \lambda'^*_\gamma - \lambda'_\beta}} \left[\Phi_{\lambda_\alpha} \lambda'_\alpha \right] . \quad (2.79)$$

One might ask what happens if we represent e.g. $V_{\Delta(\beta)}$ by any of the $V_{\tilde{\beta}}$ with $\tilde{\beta}$ defined by eq. (2.37). In fact, this does not change the fusion rules. Indeed, as discussed below eq. (2.39), one has $\tilde{\lambda}_\beta = \lambda_\beta - r_\beta$ and $\tilde{\lambda}'_\beta = \lambda'_\beta - r'_\beta$ with r_β and r'_β positive or zero roots. Thus eqs. (2.73) become $\lambda_\alpha = \lambda_\beta + \lambda_\gamma - R$, $\lambda'_\alpha = \lambda'_\beta + \lambda'_\gamma - R'$ with $R = r + r_\beta$, $R' = r' + r'_\beta$ again positive or zero roots.

Consider the example of W_3 . Then the λ 's are highest weights of $su(3)$, see the figure given above. To simplify the discussion assume $\lambda'_\beta = \lambda'_\gamma = 0$. Then the fusion rules (2.79) tell us the obvious result that $\lambda'_\alpha = 0$, too. Next suppose that $\lambda_\beta = \lambda_\gamma = \lambda_1$, the highest weight of the 3-representation. Since $\lambda_1^* = \lambda_2$, the highest weight of the $\bar{3}$ -representation, the fusion rules imply $\lambda_\alpha \leq 2\lambda_1$ and

$\lambda_\alpha^* \geq \lambda_2 - \lambda_1$. The first condition has solutions $\lambda_\alpha = 2\lambda_1$ and λ_2 whereas the second condition has solutions $\lambda_\alpha^* = 2\lambda_2, \lambda_1, \dots$, i.e. $\lambda_\alpha = 2\lambda_1, \lambda_2, \dots$. The common solutions to both conditions are $\lambda_\alpha = 2\lambda_1$ or λ_2 which are the highest weights of the 6 and the $\bar{3}$ -representations respectively. (Of course, $2\lambda_1$ was to be expected from eq. (2.68), but that simple argument did not yield λ_2). You will recognize that the fusion rules follow the pattern of composition of $su(3)$ -representations: $3 \times 3 = 6 + \bar{3}$! This is true not only for this example but in general.

2.4 MINIMAL MODELS

The fusion rules just described implement a truncation "from below": fusion always leads to λ 's that are $\lambda = \sum_i (l_i - 1)\lambda_i$ with $l_i - 1 \geq 0$ (and similarly for the λ'). However, there is no truncation "from above"; there are no maximal l_i , and the operator product algebra involves infinitely many primary fields.

Such a truncation "from above" happens if $\alpha_+^2 = \frac{p'}{p}$ (hence $\alpha_-^2 = \frac{p}{p'}$) is a rational number (p and p' are assumed relatively prime). One then has

$$\alpha_- = -\frac{p}{\sqrt{pp'}}, \quad \alpha_+ = \frac{p'}{\sqrt{pp'}}, \quad 2\alpha_0 = \frac{p' - p}{\sqrt{pp'}} \quad (2.80)$$

and from eq. (2.66)

$$\alpha - 2\alpha_0\rho = \frac{1}{\sqrt{pp'}} \sum_{i=1}^{n-1} (pl_i - p'l'_i) \lambda_i. \quad (2.81)$$

Thus only the combinations $pl_i - p'l'_i$ are relevant. Changing $l_i \rightarrow l_i + rp'$ and $l'_i \rightarrow l'_i + rp$ (r an integer) does not change α , and hence corresponds to the same highest weight. As a consequence the corresponding vertex operators must be identified. There are many more identifications to be performed, in particular changing

$$\lambda \rightarrow (p' - 2)\rho - \lambda^*, \quad \lambda' \rightarrow (p - 2)\rho - \lambda'^* \quad (2.82)$$

corresponds to $\alpha = -\alpha_- \lambda - \alpha_+ \lambda' \rightarrow -(p'\alpha_- + p\alpha_+)\rho + 2(\alpha_- + \alpha_+)\rho + \alpha_- \lambda^* + \alpha_+ \lambda'^* = 4\alpha_0\rho - \alpha^*$, which is a symmetry of the Δ_k .

Now one can compute again the fusion rules but using the representation (2.82) instead. This results in the conditions

$$\lambda_\alpha \geq \lambda_\beta + \lambda_\gamma - (p' - 2)\rho, \quad \lambda_\alpha \leq \lambda_\beta - \lambda_\gamma^* + (p' - 2)\rho, \quad \lambda_\alpha \leq \lambda_\gamma - \lambda_\beta^* + (p' - 2)\rho \quad (2.83)$$

where we also used that $\rho^* = \rho$ and that the inequalities are preserved by the star operation (since it permutes the simple roots e_i and thus also the positive roots among themselves). This will now lead to truncation from above. Imposing both fusion rules simultaneously we obtain

$$\left[\Phi_{\lambda_\beta \lambda'_\beta} \right] \times \left[\Phi_{\lambda_\gamma \lambda'_\gamma} \right] = \sum_{\substack{\lambda_\beta + \lambda_\gamma - (p' - 2)\rho \leq \lambda_\alpha \leq \lambda_\beta + \lambda_\gamma \\ \lambda_\beta - \lambda_\gamma^* \leq \lambda_\alpha \leq \lambda_\beta - \lambda_\gamma^* + (p' - 2)\rho \\ \lambda_\gamma - \lambda_\beta^* \leq \lambda_\alpha \leq \lambda_\gamma - \lambda_\beta^* + (p' - 2)\rho}} \sum_{\substack{\lambda'_\beta + \lambda'_\gamma - (p - 2)\rho \leq \lambda'_\alpha \leq \lambda'_\beta + \lambda'_\gamma \\ \lambda'_\beta - \lambda'_\gamma^* \leq \lambda'_\alpha \leq \lambda'_\beta - \lambda'_\gamma^* + (p - 2)\rho \\ \lambda'_\gamma - \lambda'_\beta^* \leq \lambda'_\alpha \leq \lambda'_\gamma - \lambda'_\beta^* + (p - 2)\rho}} \left[\Phi_{\lambda_\alpha \lambda'_\alpha} \right]. \quad (2.84)$$

In fact, we have not used all possible field identifications and one can find even more severe selection rules. It turns out that the operator product algebra closes on a finite number of primary fields:

$$\mathcal{A} = \bigoplus_{\sum l_i \leq p' - 1} \bigoplus_{\sum l'_i \leq p - 1} \Phi_{\lambda(l), \lambda'(l')}. \quad (2.85)$$

These models are called minimal models. Note that the (incomplete) fusion rules (2.84) would have allowed a slightly bigger set.

Since $2\alpha_0 = \alpha_- + \alpha_+ = \frac{p' - p}{\sqrt{pp'}}$ the central charges of these (p, p') minimal models are

$$c_{pp'}^n = (n - 1) \left(1 - n(n + 1) \frac{(p - p')^2}{pp'} \right). \quad (2.86)$$

An example: the 3-state Potts model

Rather than giving the exact fusion rules and deriving eq. (2.85) in general, we think it is more instructive to consider an example: the 3-state Potts model. Take $n = 3$, $p = 4$, $p' = 5$. Then the central charge is $c = \frac{4}{5}$. There are six W -primary fields: the identity $\mathbf{1}$, the two spin fields σ and σ^+ (with $\Delta_2 = \frac{1}{15}$), the two parafermion fields ψ and ψ^+ (with $\Delta_2 = \frac{2}{3}$) and the energy density ϵ (with

$\Delta_2 = \frac{2}{5}$). Each of them has three representatives V_α with $\alpha = -\alpha_- \lambda - \alpha_+ \lambda'$ such that λ and λ' are dominant weights. We give the list of the corresponding pairs (λ, λ') :

$$\begin{aligned}
 \mathbf{1} &: (0, 0) \simeq (2\lambda_1, \lambda_1) \simeq (2\lambda_2, \lambda_2) \\
 \sigma &: (\lambda_1, 0) \simeq (\lambda_1 + \lambda_2, \lambda_1) \simeq (\lambda_2, \lambda_2) \\
 \sigma^+ &: (\lambda_2, 0) \simeq (\lambda_1, \lambda_1) \simeq (\lambda_1 + \lambda_2, \lambda_2) \\
 \psi^+ &: (2\lambda_1, 0) \simeq (2\lambda_2, \lambda_1) \simeq (0, \lambda_2) \\
 \psi &: (2\lambda_2, 0) \simeq (0, \lambda_1) \simeq (2\lambda_1, \lambda_2) \\
 \epsilon &: (\lambda_1 + \lambda_2, 0) \simeq (\lambda_2, \lambda_1) \simeq (\lambda_1, \lambda_2) .
 \end{aligned} \tag{2.87}$$

It is then easy to compute the fusion of any two fields, e.g. $\psi \times \psi^+$. Representing $\psi \simeq (2\lambda_1, \lambda_2)$ and $\psi^+ \simeq (0, \lambda_2)$ one has by eq. (2.79)

$$\psi \times \psi^+ \simeq (2\lambda_1, \lambda_2) \times (0, \lambda_2) = (2\lambda_1, 2\lambda_2) + (2\lambda_1, \lambda_1) \simeq \omega + \mathbf{1} \tag{2.88}$$

with some new field ω . However, if one chooses to represent $\psi \simeq (2\lambda_2, 0)$ and $\psi^+ \simeq (0, \lambda_2)$ one has

$$\psi \times \psi^+ \simeq (2\lambda_2, 0) \times (0, \lambda_2) = (2\lambda_2, \lambda_2) \simeq \mathbf{1} . \tag{2.89}$$

Consistency between both equations implies

$$\psi \times \psi^+ = \mathbf{1} . \tag{2.90}$$

Fusion of ψ and ψ^+ cannot generate the new field ω . In a similar way one has

$$\begin{aligned}
 \sigma \times \sigma &\simeq (\lambda_1, 0) \times (\lambda_1, 0) = (2\lambda_1, 0) + (\lambda_2, 0) \simeq \psi^+ + \sigma^+ \\
 \psi \times \psi &\simeq (2\lambda_2, 0) \times (0, \lambda_1) = (2\lambda_2, \lambda_1) \simeq \psi^+
 \end{aligned} \tag{2.91}$$

etc. One finds that the fusion algebra closes on the six fields $\mathbf{1}, \sigma, \sigma^+, \psi, \psi^+$ and ϵ . They all satisfy $\sum l_i \leq 4 = p' - 1$ and $\sum l'_i \leq 3 = p - 1$ as claimed in the general formula (2.85).*

* We also note that the three fields $\mathbf{1}, \psi$ and ψ^+ satisfy a closed fusion sub-algebra. This is nothing but the $N = 3$ parafermion algebra [23] which is known to have central charge $c = \frac{4}{5}$ and is described by the same coset (see next subsection).

2.5 UNITARY MODELS - COSET CONSTRUCTION

This subsection is based on ref. 2. It turns out that for $p' = p + 1$ the minimal models are unitary. Indeed, for $p' = p + 1$ they coincide with the theories obtained by the coset construction based on

$$\frac{\hat{su}(n)_k \otimes \hat{su}(n)_1}{\hat{su}(n)_{k+1}} \quad , \quad p = k + n \quad (2.92)$$

Recall that for $\hat{su}(n)_k$ the generators satisfy $[J_l^a, J_m^b] = f^{abc} J_{l+m}^c + kl\delta_{l+m,0}\delta^{ab}$. The corresponding generators of $\hat{su}(n)_1$ will be denoted by $\tilde{J}_l^a$. Their sums $J_m^a + \tilde{J}_m^a$ generate the diagonal subalgebra of level $k + 1$. Recall also the Sugawara construction of the stress tensor:

$$T_k(z) = \frac{1}{2(k+n)} : \sum_a J^a(z) J^a(z) : \quad (2.93)$$

where $: \dots :$ indicates normal ordering with respect to the mode index of the current J^a . T_k generates a Virasoro algebra with central charge

$$c_k = \frac{k(n^2 - 1)}{k + n} . \quad (2.94)$$

In the same way one has

$$T_1 = \frac{1}{2(1+n)} : \sum_a \tilde{J}^a \tilde{J}^a : , \quad c_1 = \frac{(n^2 - 1)}{1 + n} = n - 1 \quad (2.95)$$

(this is equivalent to the vertex operator construction of the affine algebra in terms of $n - 1$ free fields) and also for the diagonal subalgebra

$$T_{k+1} = \frac{1}{2(k+1+n)} : \sum_a (J^a + \tilde{J}^a)(J^a + \tilde{J}^a) : , \quad c_{k+1} = \frac{(k+1)(n^2 - 1)}{k+1+n} . \quad (2.96)$$

Since $[(T_k)_l, J_m^a] = -mJ_{l+m}^a$ etc., one can easily check that

$$T = T_k + T_1 - T_{k+1} \quad (2.97)$$

commutes with $J^a + \tilde{J}^a$ and thus with T_{k+1} . Hence it is the stress tensor on the

coset and its central charge is

$$c = c_k + c_1 - c_{k+1} = (n-1) \left(1 - \frac{n(n+1)}{(k+n)(k+1+n)} \right) = c_{p,p+1}^n \quad \text{with } p = k+n. \quad (2.98)$$

We see that the central charge comes out correctly. What about the conformal dimensions? One has for the Δ_2 (Δ_2^k refers to T_k , etc.)

$$\Delta_2^k = \frac{\lambda' \cdot (\lambda' + 2\rho)}{2(k+n)}, \quad \Delta_2^1 = \frac{\epsilon \cdot (\epsilon + 2\rho)}{2(1+n)}, \quad \Delta_2^{k+1} = \frac{\lambda \cdot (\lambda + 2\rho)}{2(k+1+n)} \quad (2.99)$$

where λ', ϵ and λ are highest weights of $su(n)$, and unitarity requires

$$\lambda' \cdot \psi \leq k, \quad \lambda \cdot \psi \leq k+1, \quad \epsilon \cdot \psi \leq 1 \quad (2.100)$$

ψ being the highest root of $su(n)$ (normalized as $\psi^2 = 2$). Thus $\epsilon = 0$ or $\epsilon = \lambda_i$. One further has the obvious requirement

$$\lambda = \lambda' + \epsilon \quad \text{modulo a root}. \quad (2.101)$$

We have the following **Lemma**:

$$\Delta_2^k + \Delta_2^1 - \Delta_2^{k+1} = \Delta_2(\alpha = -\alpha_- \lambda - \alpha_+ \lambda') \quad \text{mod } \mathbf{Z} \quad (2.102)$$

(where as usual $\Delta_2(\alpha) = \frac{1}{2} \alpha \cdot (\alpha - 4\alpha_0\rho)$) provided one identifies $p = k+n$, i.e. $\alpha_-^2 = \frac{p}{p+1} = \frac{k+n}{k+n+1}$ and $\alpha_+^2 = \frac{k+n+1}{k+n}$.

Proof: First note that $\alpha_- \alpha_+ = -1$ and $\alpha_- + \alpha_+ = 2\alpha_0$ imply that $\alpha_- = -\frac{p}{\sqrt{p(p+1)}}$, $\alpha_+ = \frac{p+1}{\sqrt{p(p+1)}}$, $2\alpha_0 = \frac{1}{\sqrt{p(p+1)}}$ and thus $\alpha_- = -p2\alpha_0$, $\alpha_+ = (p+1)2\alpha_0$. Then

$$\begin{aligned} \Delta_2(\alpha) &= \frac{1}{2}(\lambda \alpha_- + \lambda' \alpha_+) \cdot (\lambda \alpha_- + \lambda' \alpha_+ + 4\alpha_0\rho) \\ &= \frac{1}{2p(p+1)}(\lambda'(p+1) - \lambda p) \cdot (\lambda'(p+1) - \lambda p + 2\rho). \end{aligned} \quad (2.103)$$

On the other hand

$$\begin{aligned}
 & \Delta_2^k - \Delta_2^{k+1} + \frac{1}{2}(\lambda - \lambda')^2 \\
 &= \frac{(p+1)\lambda' \cdot (\lambda' + 2\rho) - p\lambda \cdot (\lambda + 2\rho) + p(p+1)(\lambda^2 + \lambda'^2 - 2\lambda \cdot \lambda')}{2p(p+1)} \\
 &= \frac{(p+1)^2\lambda'^2 + p^2\lambda^2 - 2p(p+1)\lambda \cdot \lambda' + 2(p+1)\lambda' \cdot \rho - 2p\lambda \cdot \rho}{2p(p+1)} \quad (2.104) \\
 &= \frac{((p+1)\lambda' - p\lambda) \cdot ((p+1)\lambda' - p\lambda + 2\rho)}{2p(p+1)} = \Delta_2(\alpha) .
 \end{aligned}$$

Thus we find that

$$\Delta_2(\alpha) - (\Delta_2^k + \Delta_2^1 - \Delta_2^{k+1}) = -\Delta_2^1 + \frac{1}{2}(\lambda - \lambda')^2 \equiv \delta . \quad (2.105)$$

Since $\lambda = \lambda' + \epsilon - r$ (r a root) it follows that $\frac{1}{2}(\lambda - \lambda')^2 = \frac{1}{2}(\epsilon - r)^2 = \frac{1}{2}\epsilon^2 + m$ where $m = r^2/2 - \epsilon \cdot r$ is an integer. Thus $\delta = \frac{1}{2}\epsilon^2 - \frac{\epsilon \cdot (\epsilon + 2\rho)}{2(1+n)} + m$. But the quantity $n\epsilon^2 - 2\epsilon \cdot \rho$ vanishes for all allowed $\epsilon = 0, \lambda_i$ as was noted below eq. (2.14). It follows that $\delta = m$ which completes the proof.

The reason why equation (2.102) only holds modulo integers is that the W -primary fields are not necessarily obtained from the Kac Moody primary fields but may involve descendent fields.

So far we have seen how the coset stress tensor yields the correct values of the central charge and of the conformal dimensions $\Delta_2(\alpha)$ for the $(p, p+1)$ -minimal W_n -model. In ref. 2 the spin-3 generator $W^{(3)}$ was constructed in a similar way out of products of 3 currents of the form $d_{abc}J^aJ^bJ^c$, $d_{abc}\tilde{J}^a\tilde{J}^bJ^c$, etc where the d_{abc} are the completely symmetric traceless d -symbols of $su(n)$ associated with the Casimir invariant of order 3. In principle it should be possible to construct also the higher-spin generators $W^{(k)}$ for $k \geq 4$ in the same way, associated to the higher order Casimir invariants, but a look at ref. 2 tells you that this might be a rather tedious computation!

4. Lecture

The fourth lecture described a way to obtain the Ward identities of W_n (which imply all OPEs) by a certain Hamiltonian reduction of $SL(n, C)$ gauge theories in two dimensions. In fact the results are classical in the sense that only the one-contraction parts of the OPEs are obtained which means that one disregards normal-ordering effects. This approach allows us to learn a good deal about classical W -algebras. Since this is described in detail and in a hopefully rather pedagogical way in a recent preprint, I do not rephrase its content here but invite the reader to consult sections 2 and 3 of ref. 19.

REFERENCES

- [1] A.B. Zamolodchikov, *TMP* **65** (1985) 1205;
V.A. Fateev and A.B. Zamolodchikov, *Nucl. Phys.* **B280**[FS18] (1987) 644.
- [2] F. Bais, P. Bouwknegt, M. Surridge and K. Schoutens, *Nucl. Phys.* **B304** (1988) 348, 371.
- [3] J. Balog, L. Feher, P. Forgacs, L. O’Raifeartaigh and A. Wipf, *Phys. Lett.* **B227** (1989) 214, *Ann. Phys.* **203** (1990) 76.
- [4] I.M. Gelfand and I. Dorfman, *Funct. Anal. Appl.* **15** (1981) 173;
L.A. Dikii, *Comm. Math. Phys.* **87** (1982) 127.
- [5] V.G. Drinfeld and V.V. Sokolov, *Sov. Math.* **30** (1985) 1975.
- [6] J.-L. Gervais, *Phys. Lett.* **160B** (1985) 277.
- [7] K. Yamagishi, *Phys. Lett.* **B205** (1988) 466;
P. Mathieu, *Phys. Lett.* **B208** (1988) 101;
I. Bakas, *Phys. Lett.* **B213** (1988) 313.
- [8] P. Di Francesco, C. Itzykson and J.-B. Zuber, "Classical W -algebras" Princeton/Saclay preprint PUPT-1211/S.Ph-T/90-149 (October 1990).
- [9] A. Bilal and J.-L. Gervais, *Phys. Lett.* **B206** (1988) 412, *Nucl. Phys.* **B314** (1989) 646, *Nucl. Phys.* **B318** (1989) 579.
- [10] M. Jimbo, A. Kuniba, T. Miwa and M. Okado, *Comm. Math. Phys.* **119** (1988) 543.

- [11] A. Bilal, Nucl. Phys. **B330** (1990) 399, Int. J. Mod. Phys. **A5** (1990) 1881.
- [12] M. Fukuma, H. Kawai and R. Nakayama, Int. J. Mod. Phys. **A6** (1991) 1385, and "Infinite dimensional Grassmannian structure of two-dimensional quantum gravity", Tokyo preprint TU-572 (November 1990);
R. Dijkgraaf, E. Verlinde and H. Verlinde, Nucl. Phys. **B348** (1991) 435;
J. Goeree, "W-constraints in 2D gravity", Utrecht preprint THU-19 (September 1990).
- [13] E. Bergshoeff, B. De Wit and M. Vasiliev, "The structure of the super $W_\infty(\lambda)$ -algebra", CERN preprint CERN-TH. 6021/91 (March 1991), and references therein.
- [14] A. Bilal, Phys. Lett. **B238** (1990) 239;
C. Ahn, K. Schoutens and A. Sevrin, "The full structure of the super W_3 algebra", Stony Brook preprint ITP.SB-90-66 (September 1990);
K. Schoutens and A. Sevrin, "Minimal super W_N algebras in coset conformal field theories", CERN-preprint CERN-TH.5845/90 (November 1990).
- [15] B.L. Feigin and D.B. Fuchs, Funct. Anal. Appl. **16** (1982) 114, **17** (1983) 241.
- [16] V.A. Fateev and S.L. Lykhanov, Int. J. Mod. Phys. **A3** (1988) 507.
- [17] S.L. Lykhanov and V.A. Fateev, "Additional symmetries and exactly soluble models in two-dimensional conformal field theory: I, II and III", Landau Institute preprints 1988 and 1989.
- [18] R. Gilmore, "Lie groups, Lie algebras and some of their applications", John Wiley & Sons, New York, 1974.
- [19] A. Bilal, V.V. Fock and I.I. Kogan, "On the origin of W -algebras", CERN preprint CERN-TH.5965/90 (December 1990), to appear in Nucl. Phys. B.
- [20] S.L. Lykhanov, Funct. Anal. Appl. **22** no 4 (1988) 1.
- [21] A.A. Belavin, A.M. Polyakov and A.B. Zamolodchikov, Nucl. Phys. **B241** (1984) 333.
- [22] V.I. Dotsenko and V.A. Fateev, Nucl. Phys. **B240**[FS12] (1984) 312.
- [23] A.B. Zamolodchikov and V.A. Fateev, JETP **62** (1985) 215.

DUALITY INVARIANT EFFECTIVE STRING ACTIONS AND AUTOMORPHIC FUNCTIONS FOR (2, 2) STRING COMPACTIFICATIONS

Dieter Lüst

CERN, CH-1211 Geneva 23, Switzerland

Abstract

We discuss the constraints of target-space duality symmetries on the effective action of string compactifications. First, taking toroidal/orbifold compactifications as an explicit example we show the relevance of modular functions for the construction of the (non-perturbative) effective $N = 1$ superpotential as well as for the target-space free energy of string compactifications. Finally, in the context of general Calabi-Yau compactifications, we define automorphic functions in terms of holomorphic sections of a line bundle over the Calabi-Yau moduli space. This construction naturally generalizes the Dedekind function of the previously discussed multitoroidal compactifications.

1. Introduction

It is undoubtedly a very important aim to establish a clearer understanding about the relation between four-dimensional heterotic string theories and low energy physics. Of especial “phenomenological” interest are those four-dimensional string vacua with $N = 1$ space-time supersymmetry. A large subclass of them can be obtained via compactification of the ten-dimensional heterotic string [1] on a six-dimensional Calabi-Yau ($C - Y$) space [2] which can be also described by an internal (2,2) superconformal field theory (*SCFT*) with $c = 22$, $\bar{c} = 9$. However infinitely many of these models are continuously connected by deformations of the underlying *SCFT*. These deformations are described by the so-called moduli parameters T_i which are local coordinates of the moduli space $\mathcal{M}$, $T_i \in \mathcal{M}$. The

moduli are just the coupling constants of the marginal operators of the *SCFT*. Seen from the compactification point of view, the moduli correspond to the background parameters of the six-dimensional $C - Y$ space. Finally, in the corresponding four-dimensional effective field theory each modulus leads to a massless scalar field $T_i(x)$. In any order of string perturbation theory the effective scalar potential of the scalars T_i is completely flat; the arbitrary vacuum expectation values (*vev*'s) of these fields exactly determine the $C - Y$ background parameters.

In addition to the internal moduli fields there exists another universal complex scalar with vanishing potential, the so-called S -field, whose real part is the dilaton field with *vev* determining the string coupling constant (the loop counting parameter) and whose imaginary part is an axion field: $S = \frac{1}{g^2} + i\theta$.

This kind of scenario causes some immediate “phenomenological” problems. First, the flatness of the moduli potential implies that $N = 1$ space-time supersymmetry is left unbroken in any order of string perturbation theory. Second, due to the huge vacuum degeneracy the theory suffers from a lack of predictivity, since many coupling constants like gauge or Yukawa couplings are moduli (field) dependent quantities and therefore undetermined: $g^2 = g^2(S, T_i)$, $h = h(S, T_i)$. Moreover, for special values of the moduli there may be enhanced gauge symmetries, associated to vector bosons which become massless at these special points. Thus, for all of these reasons it is crucial to get a proper understanding of the dynamics which eventually determine the *vev*'s of the T_i , S fields and, at the same time, provide a mass for these fields.

Specifically, one expects that non-perturbative (∞ -genus) string effects will lift the vacuum degeneracy. Unfortunately, despite many efforts, a non-perturbative formulation of string theory is not available at the present moment, and these fundamental questions cannot be analyzed from first principles. Nevertheless, very useful information about non-perturbative effects can be obtained from the low energy effective field theory of the light string degrees of freedom which is obtained after integrating out all massive string modes. It is very important to realize that this effective action is highly constrained by the so-called target space duality symmetries [3-5]. These are entirely stringy symmetries which act as discrete reparametrizations on the moduli, $T_i \rightarrow \tilde{T}_i(T_i)$, but leave the underlying *SCFT* invariant. The simplest example is the $R \rightarrow \frac{1}{R}$ duality symmetry for

circle compactification where R is the radius of the internal circle. For toroidal (orbifold) compactifications the target space duality group Γ always contains the modular group $PSL(2, \mathbf{Z})$. As shown in [6], the modular invariance of the effective action establishes a connection between the effective $N = 1$ superpotential and the theory of modular functions. Finally, for general (2,2) $C - Y$ compactifications target space duality symmetries, which in general do not contain the modular group, seem also to be present, as shown for particular cases in [7]. Thus the $N = 1$ supersymmetric effective action is in general determined by the automorphic functions of the target space duality group Γ .

In the next section we briefly review the $R \rightarrow \frac{1}{R}$ duality symmetry, the target space modular invariance and its implications on the effective action of the moduli fields. In section three we define the topological target space free energy and its relation to the non-perturbative superpotential in the context of the duality invariant gaugino condensation mechanism. Finally, in section four, we present the construction of automorphic functions for general (2,2) compactifications.

2. Target Space Modular Invariance

2.1 CIRCLE COMPACTIFICATION

Let us briefly recall the duality symmetry for closed string compactification on a one-dimensional circle S_1 [3]. The periodicity condition for a closed string on a circle has the form

$$X(\sigma + 2\pi, \tau) = X(\sigma, \tau) + 2\pi n R. \quad (2.1)$$

The second term describes string states which are only closed on the circle with radius R , however not on the real axis. These states correspond to the so-called winding states, and they are labelled by their winding number n . This phenomenon has no field theoretical counterpart, the winding states are of purely stringy nature. The mass of a compactified string takes the form (we set the string tension $\alpha' = 2$):

$$M^2 = \frac{m^2}{R^2} + \frac{1}{4}n^2 R^2 + N_L + N_R - 2. \quad (2.2)$$

The term m/R^2 is the contribution from the quantized momenta in the compact dimension, $m \in \mathbf{Z}$ is the momentum number of a state; the term $\frac{1}{4}n^2 R^2$ is the en-

ergy required to wrap the string around the circle and N_L (N_R) are the left- (right-) moving oscillator numbers. Physical states have to satisfy the reparametrization constraint $N_R - N_L = mn$. Thus the states with no oscillator excitations must satisfy $mn = 0$, i.e. they are either purely momentum or winding. We will call these states which are entirely due to the compactification *topological* states. Note that the topological spectrum does not contain those four states which become massless at $R = \sqrt{2}$ and lead to an enhanced $SU(2)_L \times SU(2)_R$ Kac-Moody algebra since they have $m = \pm n = \pm 1$, $N_L(N_R) = 1$.

Inspection of the mass formula (2.2) immediately shows that the spectrum is invariant under the $\mathbf{Z}_2$ duality symmetry

$$R \rightarrow \frac{2}{R} \quad (2.3)$$

if one simultaneously exchanges also the role of the momentum and winding numbers: $m \leftrightarrow n$. Thus the duality transformation exchanges an infinite number of massive momentum and winding states. Moreover one can show that not only the string spectrum is duality invariant, but also all vertex operators and therefore all string interactions. This means that $R \rightarrow 2/R$ is a perfect automorphism of the underlying $c = 1$ CFT. It follows that the moduli space of the $c = 1$ CFT is not given by the whole positive real axis but only by the half-line $R \in \mathcal{M} = \mathbf{R}/\mathbf{Z}_2 = [\sqrt{2}, \infty)$.

2.2 TOROIDAL/ORBIFOLD COMPACTIFICATION

Consider now the compactification of a closed string on a two-dimensional torus T_2 , which is defined by two basis vectors $\vec{e}_1, \vec{e}_2$. The toroidal compactification [8] is described by four background parameters, three metric degrees of freedom $G_{ij} = \vec{e}_i \cdot \vec{e}_j$ and one antisymmetric tensor field $B_{ij} = B\epsilon_{ij}$ ($i, j = 1, 2$). It is very useful [4] to combine these four real background parameters into two complex fields like

$$T = \sqrt{G} + iB, \quad U = \frac{\sqrt{G}}{G_{22}} + i\frac{G_{12}}{G_{22}}. \quad (2.4)$$

The real part of the T -field is given by the volume of the torus (the overall “radius”), where the U -field is nothing other than the complex structure of T_2 . In

analogy to eq.(2.2) the mass formula now looks like

$$M^2 = \frac{1}{8\text{Re}T\text{Re}U} \{ |m + nTU + i(Um' + Tn')|^2 + |m + n\bar{T}U + i(Um' + \bar{T}n')|^2 \} + N_L + N_R. \quad (2.5)$$

Restricting ourselves again to the topological spectrum with $N_L = N_R = 0$, which is equivalent to $mn + m'n' = 0$, the mass formula becomes (this excludes again states which lead to an enhanced Kac-Moody algebra $SU(2)_L^2 \times SU(2)_R^2$ respectively $SU(3)_L \times SU(3)_R$):

$$M^2 = \frac{|m + nTU + i(Um' + Tn')|^2}{(T + \bar{T})(U + \bar{U})}. \quad (2.6)$$

Note that M^2 is now nicely given, up to $\text{Re}T$, $\text{Re}U$, by the modulus square of a holomorphic function of T and U . The field theoretical interpretation of this observation will be given in the next section.

The above expression (2.6) is duality and shifting invariant:

$$\begin{aligned} m' &\leftrightarrow n' & T &\leftrightarrow U, \\ m &\rightarrow m - n', \quad n' \rightarrow n' & T &\rightarrow T + i, \\ m &\rightarrow n', \quad n \rightarrow m', \quad m' \rightarrow -n, \quad n' \rightarrow -m & T &\rightarrow \frac{1}{\bar{T}}. \end{aligned} \quad (2.7)$$

In addition, one has the “charge conjugation” symmetry $U \leftrightarrow \bar{U}$, $T \leftrightarrow \bar{T}$. These are the four generators of $\Gamma = [SO(2, 2, \mathbf{Z})] = [PSL(2, \mathbf{Z})_T \times PSL(2, \mathbf{Z})_U \times \mathbf{Z}_2 \times \mathbf{Z}_2]$, which define the (T, U) target-space duality transformations. The duality transformations on the T -field are of stringy nature since they induce the exchange of momentum and winding states. However the modular transformations involving the U -field are field-theoretical; they are also present setting the two winding numbers n and n' to zero in (2.6). These transformations are of geometrical origin; the modular group $PSL(2, \mathbf{Z})$ for the U -field is just that of two-dimensional gravity.

To obtain finally a six-dimensional compactification one can simply consider orbifold constructions [9] which are just the direct product of three two-dimensional tori possibly divided out by some discrete symmetry: $O_6 = T_2 \otimes T_2 \otimes T_2/P$. This

amounts to setting some or all of the complex structure moduli U_A ($A = 1, 2, 3$) to fixed values, e.g. $U_A = 1$. For further simplification let us keep only the overall breathing mode, i.e. $T = T_1 = T_2 = T_3 = R^2 + iB$. Then the duality group is given by the modular group $PSL(2, \mathbf{Z})$ acting on T as

$$T \rightarrow \frac{aT - ib}{icT + d} \quad a, b, c, d \in \mathbf{Z} \quad ad - bc = 1. \quad (2.8)$$

These transformations divide the complex half plane ($\text{Re}T > 0$) into equivalence classes or fundamental domains for which two interior points are not mapped into each other under (2.8). Thus the moduli space $\mathcal{M} = SU(1, 1)/U(1)/PSL(2, \mathbf{Z}) \simeq \mathbf{H}_+/PSL(2, \mathbf{Z})$ of the T -field can be taken as $-\frac{1}{2} \leq \text{Im}T < \frac{1}{2}$, $|T| \geq 1$.

2.3 MODULAR INVARIANT EFFECTIVE SUPERGRAVITY ACTION

Let us now discuss the target space modular invariant effective $N = 1$ supergravity action relevant for orbifold types of compactifications. For simplicity we include as low energy degrees of freedom only the two chiral superfields S, T . The $N = 1$ supergravity matter couplings are determined by the following function [10]:

$$G(S, T, \bar{S}, \bar{T}) = K(S, T, \bar{S}, \bar{T}) + \log |W(S, T)|^2. \quad (2.9)$$

For the Kähler potential $K(S, T, \bar{S}, \bar{T})$ we take its tree-level form [11]

$$K(S, T, \bar{S}, \bar{T}) = -\log(S + \bar{S}) - 3\log(T + \bar{T}). \quad (2.10)$$

Then it is not difficult to show that a target space modular transformation (2.8) acts as a Kähler transformation on the Kähler potential:

$$K \rightarrow K + \log |icT + d|^6. \quad (2.11)$$

(The tree-level S -field is invariant under modular transformations; for simplicity we assume the absence of S - T mixing in the Kähler potential.) In order to obtain a modular invariant effective supergravity action (a modular invariant scalar potential and Yukawa couplings) the G -function (2.9) must itself be modular invariant.

It follows that the superpotential $W(S, T)$ has to transform under target space modular transformations as [6]

$$W(T) \rightarrow e^{i\delta(a,b,c,d)} \frac{W(T)}{(icT + d)^3}. \quad (2.12)$$

Thus, $W(T)$ has to be a modular function of modular weight -3 .^{*} The field independent phase $\delta(a, b, c, d)$ is the so-called multiplier system. (See [13] for details.) This requirement will give strong restrictions to the form of the entirely S, T -dependent non-perturbative superpotential $W_{NP}(S, T)$ regardless of, what the actual non-perturbative mechanism responsible for the emergence of W_{NP} is. Its most general form is given by [6],[13]

$$W_{NP}(S, T) = \frac{\Omega(S)H(T)}{\eta(T)^6}. \quad (2.13)$$

$\Omega(S)$ is a so far unspecified function of S ; due to its non-perturbative (infinite genus) character one expects that $\Omega(S)$ has an expansion in terms of powers of e^S . $\eta(T)$ is the Dedekind function and $H(T)$, being an arbitrary modular invariant function with in general non-vanishing multiplier system, is a rational function of the absolute modular invariant $j(T)$. Note that eq.(2.13) implies that $W(T)$ has an expansion in $e^{T/\alpha'}$ (for large T) after reintroducing the dimensionful slope parameter α' . This form is obviously non-perturbative in the two-dimensional σ -model coupling constant α'/R^2 and clearly shows the relevance of the contribution of the world-sheet instantons [14] to $W(T)$. The behaviour of $W(T)$ will largely determine the vacuum structure of the $N = 1$ supergravity action. In particular, the scalar potential will inherit its singularities. In fact, since $W(T)$ has negative modular weight it follows that $W(T)$ will in general have singularities at ∞ and/or at finite values of T . Demanding the absence of zeroes and poles inside the fundamental region, the only choice is $H(T) = \text{const.}$ We will restrict ourselves in the following to this physically most important case. (A discussion about the physical relevance of possible poles/zeroes inside the fundamental region can be found in [13].)

* Including matter fields, one can explicitly show [12] that the tree-level cubic superpotential has indeed modular weight -3 .

To determine the non-perturbative vacuum structure of the theory we compute the scalar potential resulting from eqs.(2.10),(2.13):

$$\begin{aligned} V &= |h^S|^2 G_{SS^*}^{-1} + |h^T|^2 G_{TT^*}^{-1} - 3 \exp(G) \\ &= \frac{1}{S_R T_R^3 |\eta|^{12}} \left\{ |S_R \Omega_S - \Omega|^2 + \frac{T_R^2}{3} \left| \frac{3}{2\pi} \hat{G}_2 \right|^2 |\Omega|^2 - 3 |\Omega|^2 \right\} \end{aligned} \quad (2.14)$$

where $T_R = T + \bar{T}$, $S_R = S + \bar{S}$, $\Omega_S = d\Omega/dS$ and $\hat{G}_2$ is the non-holomorphic Eisenstein function of modular weight two. Space-time supersymmetry is spontaneously broken if at the minimum of V one of the auxiliary fields

$$h^i = \exp\left(\frac{1}{2}G\right) G^i = |W| \exp\left(\frac{1}{2}K\right) \left(K^i + \frac{W^i}{W}\right) \quad (2.15)$$

has a non-zero vacuum expectation value. The analysis of the extrema of V [15],[13] yields the following main results:

(i) $h^S = 0$ is always an extremum. In fact one can show that the conditions for having a minimum for large $\text{Re}S$ are easy to fulfil for this type of extrema. An example is $\Omega = c + e^{\alpha S}$, allowing for small c a solution for $S_R \Omega_S - \Omega$ at weak coupling.

(ii) Compactification is dynamically preferred. The minimum of V occurs at $T_{\min} < \infty$. In appropriate units, $\text{Re}T_{\min} \sim 1.2$, $\text{Im}T = 0$. Note that at the minimum $\text{Im}T$ is zero (modulo integers) having interesting consequences for the strong CP problem [16]. At this minimum $N = 1$ local supersymmetry is spontaneously broken in the T -field sector since $h^T \neq 0$. For large $\text{Re}T$ the scalar potential (more precisely $|W|^2$) exponentially diverges like $\exp \pi T$. In the following section we will explain why one expects this divergence of W , V on general physical grounds.

3. The Topological Free Energy of $N = 1$ Orbifolds

3.1 THE TARGET SPACE FREE ENERGY

Here we want to define a very useful topological quantity [17], namely the free energy F of the compactified heterotic string. This always appears when computing amplitudes with light states as external legs, and all massive states entering in loops. Therefore the free energy F must be a target-space duality invariant function of the moduli fields. In the field theory context, the free energy F is obtained by functionally integrating out all massive states. We call the free energy a topological quantity because the path integral should be performed only over those massive states which are due to the compactification, like Kaluza-Klein and winding modes, but not over states which are also present in the uncompactified theory, like all massive oscillator states.* As we have shown for orbifold compactification, the omission of the oscillator states means that only fields with chiral masses (cf. eq.(2.6)) originating from a superpotential will contribute to F . On the other hand, non-chiral D -term masses which are responsible for the stringy Higgs effect [19], i.e. the emergence of additional massless gauge bosons at some points in the moduli space, will not contribute to F . Specifically, integration over the massive fermions leads to

$$F = \det \frac{|W_{ij}|^2}{Y}. \quad (3.1)$$

The determinant of the chiral masses, $\det W_{ij}$, is the second derivative of the superpotential of the massive states and must be an automorphic function of the duality group Γ . The factor Y in (3.1) is related to the non-canonical kinetic energy of the massive fields.

For toroidal/orbifold compactifications F can be explicitly computed from the effective field theory of the massive momentum and winding states using the results of [20]. Applying this method one obtains [17] for the free energy of the two-torus

$$F = \sum_{m,m',n,n'} \log \frac{|m + nTU + i(m'U + n'T)|^2}{(T + \bar{T})(U + \bar{U})}. \quad (3.2)$$

Note that the argument of the logarithm exactly agrees with the string mass

* This means that F is given in terms of a topological index, the Ray-Singer torsion [18].

formula (2.6). This expression has already appeared in different places in the literature. For instance, the authors of ref. [21] obtained exactly this expression when calculating one-loop corrections to gauge coupling constants. Also in the recent work of ref. [22], F is the free energy of compactified $N = 2$ critical strings.

Since F is infinite due to the infinite lattice sum, one needs a regularization procedure which respects the target space duality symmetry $\Gamma = PSL(2, \mathbf{Z})_T \times PSL(2, \mathbf{Z})_U \times \mathbf{Z}_2 \times \mathbf{Z}_2$. Since we have defined F to be a topological quantity, the sum in the above formula does not run over oscillator excitations and is therefore subject to the constraint $mn + m'n' = 0$. This constraint on the summation over n, m, n', m' can be solved in two inequivalent orbits, namely $n = m' = 0$ and $n = n' = 0$. Therefore, the free energy is just the sum of two orbits which can be evaluated using ζ -function regularization (see [17]). One obtains

$$F \Big|_{\text{reg.}} = \log[|\eta(T)|^4 |\eta(U)|^4 (T + \bar{T})(U + \bar{U})]. \quad (3.3)$$

F diverges exactly at the cusp points of the moduli space, $T, U = 0, \infty$, which correspond to the degeneration points of the spectrum. For example, for $T = \infty$ ($T = 0$) the Kaluza-Klein (winding) modes become massless and thus give an infinite contribution to the free energy.

Now consider the direct product of three two-tori keeping only a single modulus $T = T_1 = T_2 = T_3$. Then the free energy looks like

$$F \Big|_{\text{reg.}} = \left[\sum_{n,m} \log \frac{|m + inT|^6}{(T + \bar{T})^3} \right]_{\text{reg.}} = \log[|\eta(T)|^{12} (T + \bar{T})^3]. \quad (3.4)$$

We recognize that the form of the orbifold free energy is completely analogous to the form of the G -function (2.9) itself. Specifically, the T -dependent non-perturbative superpotential looks exactly like the determinant of the chiral masses which are due to the compactification:

$$W_{NP}(T) = \frac{1}{\eta(T)^6} = (\det M)_{\text{reg.}}^{-1} = \left[\prod_{m,n} (m + inT)^{-3} \right]_{\text{reg.}}. \quad (3.5)$$

$W_{NP}(T)$ diverges for large T since in this limit all Kaluza-Klein states with non-vanishing m become massless and give a infinite contribution to the sum. Anal-

ogously, for small T the infinite number of light winding states cause the same divergence.

3.2 MODULAR INVARIANT GAUGINO CONDENSATION

We will now show that the modular invariant gaugino condensation mechanism [15], [23],[13] exactly leads to the discussed type of non-perturbative superpotentials. This will also explain the reason for the identification (3.5) of $W_{NP}(T)$ with $(\det M)^{-1}$. As is well known [24], local $N = 1$ supersymmetry can be spontaneously broken by the dynamical formation of non-vanishing gaugino condensates in a hidden gauge group G ($G = E_8$). For simplicity we will assume that G is a pure Yang-Mills gauge group; the inclusion of (condensing) matter was discussed in [25]. The size of the gaugino bilinear is given by $\langle \lambda\lambda \rangle = e^{\frac{3}{2b}f(S,T)}$, where b is the $N = 1$ β -function coefficient and g^2 the modular invariant gauge coupling constant at the string scale. As recently clarified in [26] the non-perturbative superpotential is determined by the “holomorphic” gaugino condensate:

$$W_{NP}(S, T) = \langle \lambda\lambda \rangle_{\text{hol.}} = e^{\frac{3}{2b}f(S,T)}. \quad (3.6)$$

Here $f(S, T)$ is the holomorphic part of the gauge coupling constant:

$$\frac{1}{g^2} = \frac{1}{2}(f(S, T) + \bar{f}(\bar{S}, \bar{T})) - b \log(T + \bar{T}). \quad (3.7)$$

The non-holomorphic term is related [27] to σ -model anomalies. $f(S, T)$ has the form

$$f(S, T) = S + \frac{2}{3}b \log \Delta(T). \quad (3.8)$$

The T -dependent second term is the one-loop threshold correction [28] due to the intergration over the massive modes. It follows that $\Delta(T)$ is proportional to $\det M(T)$. Requiring [15], [29] modular invariance of the gauge coupling (3.7) (and the absence of zeroes and poles inside the fundamental region) $\Delta(T)$ is uniquely determined to be

$$\Delta(T) = \det M(T) = \eta(T)^{-6}. \quad (3.9)$$

The result eqs.(3.8),(3.9) has to be compared with the direct string computation [21] of the threshold correction for orbifold compactifications. Specifically, in [21]

the difference between the threshold corrections of two different gauge groups is calculated. This has the effect that the corresponding coefficient in eqs.(3.7),(3.8) is not identical to the $N = 1$ β -function coefficient. However considerations [27] on σ -model anomalies strongly indicate that $b_{N=1}$ is the correct coefficient in (3.7), (3.8) for pure Yang-Mills gauge groups.

Thus, finally putting eqs.(3.6),(3.8) and (3.9) together we obtain $W_{NP}(S, T) = e^{\frac{3}{26}S} / \det M(T) = e^{\frac{3}{26}S} / \eta(T)^6$ in agreement with our previous discussion.

4. Automorphic Functions for General Calabi-Yau Compactifications

Now let us discuss general (2,2) Calabi-Yau compactifications and consider the following function of the moduli fields T_i ($i = 1, \dots, n$):

$$\frac{|\Delta(T_i)|^2}{Y}, \quad (4.1)$$

Y is given by the Kähler potential of the moduli, $Y(T_i, \bar{T}_i) = e^{-K(T_i, \bar{T}_i)}$, where $\Delta(T_i)$ is a holomorphic function of the moduli fields T_i . Thus, $\Delta(T_i)$ provides an expression for the non-perturbative, purely moduli dependent superpotential, and we will also give some arguments (for a more exhaustive discussion see [17]) for the conjecture that $\Delta(T_i)$ is again related to the determinant of all chiral masses in the Calabi-Yau compactification spectrum. $\Delta(T_i)$ is an automorphic function of the discrete target-space duality group Γ . More precisely, it is inversely proportional to some power of the cusp form (the generalization of $\eta(T)^{24}$) of the duality group, since only for infinite (zero) radius of the internal space does the contribution of the Kaluza-Klein (winding) modes create a pole for Δ .

More generally, $\Delta(T_i)$ can be regarded as a holomorphic section of a line bundle over the moduli space $\mathcal{M}$. In particular, since the target space duality transformations $\Gamma : T_i \rightarrow \tilde{T}_i(T_j)$ act as Kähler transformations on the Kähler potential $K = -\log Y$,

$$Y \rightarrow Y |e^{-\Lambda(T_i)}|^2 \quad \text{for} \quad T_i \rightarrow \tilde{T}_i, \quad (4.2)$$

the holomorphic section Δ has to transform as

$$\Delta(T_i) \rightarrow \Delta(T_i) e^{-\Lambda(T_i)}. \quad (4.3)$$

We now construct [17] the holomorphic section $\Delta(T_i)$ for all (2,2) symmetric Calabi-Yau compactifications. For this purpose we will use the important observation [30] that the metric of the moduli space of Calabi-Yau manifolds should not only be Kähler, as implied by $N = 1$ space-time supersymmetry, but also “special Kähler”, as implied by $N = 2$ space-time supersymmetry when one regards the (2,2) theories as possible vacua for type II superstrings. Specifically, in type IIA theories the (1,1) moduli belong to vector multiplets, whereas the (2,1) moduli belong to hypermultiplets. In type IIB theories the situation is reversed. Since the particular form of the moduli space does not depend on whether we deal with a heterotic, type IIA or type IIB theory, the couplings of the moduli fields are strongly constrained by the possible couplings of $N = 2$ vector multiplets.

In the $N = 2$ superconformal tensor calculus [31], one describes the couplings of n physical vector multiplets by first introducing $n + 1$ vector multiplets, whose scalar components we denote by X^I , $I = 0, 1, \dots, n$. The X^I are holomorphic functions (sections) of the moduli: $X^I = X^I(T^i)$. The special Kähler potential can be written as

$$K = -\log(X^I \bar{\mathcal{F}}_I + \bar{X}^I \mathcal{F}_I). \quad (4.4)$$

$\mathcal{F}(X)$ is a homogeneous holomorphic function of degree 2 in the $n + 1$ variables X . This formula is independent of a particular coordinate system. However, to compare with the previous formulas it is very convenient to choose a “special” coordinate system, the “special” gauge, in the following way:

$$T^i = -iX^i/X^0. \quad (4.5)$$

Since the function $\frac{|\Delta|^2}{Y}$ must be duality-invariant, $\Delta(T^i)$ must be a holomorphic section of holomorphic degree one (Y has degree (1,1)). Holomorphic sections of exactly this degree are X^I and $\mathcal{F}_I$. Therefore the most natural ansatz for Δ is to take a product of all possible linear combinations of these two holomorphic functions [17]:

$$\log \frac{|\Delta|^2}{Y} = \left[- \sum_{M_I, N^I} \log \frac{|M_I X^I + i N^I \mathcal{F}_I|^2}{X^I \bar{\mathcal{F}}_I + \bar{X}^I \mathcal{F}_I} \right]_{\text{reg.}}. \quad (4.6)$$

As already discussed, the holomorphic section Δ is an automorphic function,

inversely proportional to the cusp form, of the Calabi–Yau duality group Γ . Δ has to transform under target space duality transformations in such a way that $\frac{|\Delta|^2}{Y}$ is a completely invariant function. This requirement implies that the product over M_I and N^I in eq.(4.6) does not go over all but only over a restricted set of integers. This restriction is equivalent to the topological level matching condition $\bar{p}_L^2 - \bar{p}_R^2 = 0$ for orbifold compactifications. To understand this, recall (see for example [32]) that the target space duality transformations act (up to a holomorphic function) as particular $Sp(2n+2)$ transformations on the vector $(X^I, i\mathcal{F}_I)$:

$$\begin{pmatrix} X^I \\ i\mathcal{F}_I \end{pmatrix} \rightarrow \begin{pmatrix} B & D \\ C & A \end{pmatrix} \begin{pmatrix} X^I \\ i\mathcal{F}_I \end{pmatrix}. \quad (4.7)$$

A, B, C, D are real constant matrices satisfying symplectic conditions and provide a linear realization of the target space modular group, i.e. $\Gamma \subset Sp(2n+2)$. Note that not any symplectic transformation corresponds to a duality transformation; the specific form of the matrices A, B, C, D depends on the particular holomorphic function $\mathcal{F}$ and on the relevant duality group. It is now evident that eq.(4.6) provides a duality invariant expression only if the symplectic transformations on X^I and $i\mathcal{F}_I$ are accompanied by the following transformations on the integers M_I and N^I :

$$\begin{pmatrix} M_I \\ -N^I \end{pmatrix} \rightarrow \begin{pmatrix} C & A \\ B & D \end{pmatrix} \begin{pmatrix} M_I \\ -N^I \end{pmatrix}. \quad (4.8)$$

This means that (M_I, N^I) must build proper representations of the symplectic duality transformations, i.e. they form a restricted set of integers. In other words, take a given pair of (M_I, N^I) ; then the action of the symplectic matrices A, B, C, D , which is in general of infinite order, generates all possible integers N_I and M^I which may contribute to Δ . This means that the product in (4.6) has to go over the specific orbits $\mathcal{O} (M^I, N_I \in \mathcal{O})$ which are generated by the particular symplectic transformations.

To obtain a clearer understanding about the generalized momentum (Kaluza-Klein) and winding spectrum of a Calabi-Yau compactification let us consider the large radius, i.e field theory, limit (the weak coupling limit of the corresponding σ -model) of the Calabi-Yau compactification. For the (1,1) moduli (the deformations of the Kähler class), $Y_{(1,1)}$ just becomes in this limit the volume of the

six-dimensional Calabi-Yau space,

$$Y_{(1,1)} = \int J \wedge J \wedge J = d_{ijk}(T_i + \bar{T}_i)(T_j + \bar{T}_j)(T_k + \bar{T}_k), \quad (4.9)$$

where $J = \sum_{i=1}^n (T^i + \bar{T}^i) V_i$, V_i being the harmonic $(1,1)$ forms in $H^{(1,1)}$. In the field theory limit d_{ijk} coincide with the intersection matrices of V_i : $d_{ijk} = \int V_i \wedge V_j \wedge V_k$. For the case of a single modulus the Kähler potential becomes $K = -3 \log(T + \bar{T})$ in the large radius limit. Now choose the special gauge eq.(4.5) and consider those "masses" in eq.(4.6) which have only non-vanishing M_0 :

$$M_{M_0}^2 = \frac{M_0^2}{Y} = \frac{M_0^2}{\int J \wedge J \wedge J}. \quad (4.10)$$

We recognize that in the field theory limit the masses $M_{M_0}^2$ are inversely proportional to the volume of the Calabi-Yau space. This fact strongly suggests that $M_{M_0}^2$ are just the ordinary field-theoretical Kaluza-Klein masses, i.e. that the momentum spectrum of the Calabi-Yau compactification is given by the states with only non-vanishing M_0 . On the other hand, for small radii, non-perturbative effects, i.e. instanton configurations modify eqs.(4.9),(4.10). At the same time the generalized winding states become light. The symplectic duality invariance forces one to include also the winding states with non-vanishing M_i and N^I in the target space free energy. Thus, the masses of the winding states seem to be determined by $M_i T^i$ and in particular by $N^I \mathcal{F}_I$. In summary, just like for toroidal/orbifold compactifications we can regard the states with non-vanishing M_0 as momentum states, and the states with non-vanishing M_i and N^I as generalized winding states (vortex configurations). Thus, the M_I and N^I define, in some sense, a generalized Calabi-Yau lattice.

Now we show that eq.(4.6) reproduces the results for the free energy as computed in the previous section for the case of the orbifold compactifications. We restrict ourselves to considering the simplest case with a single overall modulus $T = -iX^1/X^0$. The corresponding holomorphic function $\mathcal{F}(X)$ looks like $\mathcal{F} = i \frac{(X^1)^3}{X^0}$. Then the holomorphic section Δ takes the form:

$$\Delta(T) = \left[\prod_{M_I, N^I} (M_0 - iN^0 T^3 + iM_1 T + 3N^1 T^2)^{-1} \right]_{\text{reg.}}. \quad (4.11)$$

Now consider the action of the target space modular group $\Gamma = SL(2, \mathbb{Z})$. The

transformation $T \rightarrow T - i$ acts as the following symplectic transformation on M_I and N^I [33]:

$$\begin{aligned} M_0 &\rightarrow M_0 - M_1 - N^0 - 3N^1, \\ M_1 &\rightarrow M_1 + 3N^0 + 6N^1, \\ N^0 &\rightarrow N^0, \\ N^1 &\rightarrow N^0 + N^1. \end{aligned} \tag{4.12}$$

For $T \rightarrow \frac{1}{T}$ one has

$$\begin{aligned} M_0 &\rightarrow -N^0, \\ M_1 &\rightarrow -3N^1, \\ N^0 &\rightarrow M_0, \\ N^1 &\rightarrow \frac{1}{3}M_1. \end{aligned} \tag{4.13}$$

Now it is crucial to observe that these symplectic transformations define the following $SL(2, \mathbf{Z})$ orbit, i.e. the following restriction on the integers M_I and N^I :

$$\begin{aligned} M_0 &= m^3, & N^0 &= n^3, \\ M_1 &= 3m^2n, & N^1 &= -mn^2. \end{aligned} \tag{4.14}$$

m, n are now unrestricted integers, and it is easy to see that (4.11) exactly reproduces the free energy (3.4) with $\Delta(T) = [\prod_{m,n} (m + inT)^{-3}]_{\text{reg.}} = \frac{1}{\eta(T)^6}$.

5. Conclusions

In this paper we have discussed the relation between automorphic functions of discrete target-space duality groups and low-energy effective string actions which are obtained after integrating out the massive string modes. Specifically, the automorphic functions determine the non-perturbative superpotential of the $N = 1$ effective supergravity action and provide therefore important information about the non-perturbative vacuum structure of the compactified string. Secondly, the target-space free energy is also given in terms of automorphic functions. For orbifold compactifications the $N = 1$ superpotential as well as the free energy are both given in terms of the Dedekind function. This coincidence becomes clear by considering the dynamical formation of gaugino condensates and the moduli-dependent threshold corrections to the gauge coupling constants.

For arbitrary Calabi–Yau compactifications we presented the construction of the automorphic function $\Delta(T_i)$ in terms of holomorphic sections of a line bundle using the fact the Calabi–Yau moduli space is a special Kähler space. Applying this formula for the case of orbifold compactifications one rederives the Dedekind function. However it is important to mention that for general Calabi–Yau compactifications there exists in addition to $\Delta(T_i)$ another type of automorphic function which is needed to provide a duality invariant completion of the determinant of the Kähler metric of the moduli space [17].

The free energy F is a topological index, in the sense that it takes into account only those massive states which are due to the compact target space, and not those states which are also present in a flat background, like all oscillator excitations. Moreover, the general form of the automorphic functions $\Delta(T_i)$ indicates a very close relationship between Calabi–Yau compactifications and topological superconformal field theories [34], [35]. One knows [35] that $\mathcal{F}(X^I)$ is entirely determined by the couplings of the chiral primary fields [36] of the underlying $(2,2)$ $\hat{c} = 6$ superconformal field theory. However, in the topological counterparts of the superconformal models one is essentially left with the chiral ring. So we recognize that all the chiral masses $M_I X^I + i N^I \mathcal{F}_I$, and thus also the free energy of Calabi–Yau compactifications, are topological, since they are entirely determined by the $\mathcal{F}$ -function, i.e. by the topological sector of the superconformal field theory. It is a very interesting problem to further clarify this relation between Calabi–Yau compactifications and topological field theories.

ACKNOWLEDGEMENTS

I would like to thank M. Cvetič, S. Ferrara, A. Font, L. Ibáñez, C. Kounnas, F. Quevedo, A. Shapere, T. Taylor, S. Theisen, and F. Zwirner for the very fruitful collaboration on the topics presented in this talk. This work was performed as a Heisenberg fellow.

REFERENCES

- [1] D.J. Gross, J.A. Harvey, E. Martinec and R. Rohm, Phys. Rev. Lett. 54 (1985) 502; Nucl. Phys. B256 (1985) 253; Nucl. Phys. B267 (1986) 75.
- [2] P. Candelas, G. Horowitz, A. Strominger and E. Witten, Nucl. Phys. B258 (1985) 46.
- [3] K. Kikkawa and M. Yamasaki, Phys. Lett. 149B (1984) 357; N. Sakai and I. Senda, Progr. Theor. Phys. 75 (1986) 692; V.P. Nair, A. Shapere, A. Strominger and F. Wilczek, Nucl. Phys. B287 (1987) 402; B. Sathiapalan, Phys. Rev. Lett. 58 (1987) 1597.
- [4] R. Dijkgraaf, E. Verlinde and H. Verlinde, Comm. Math. Phys. 115 (1988) 649; "On moduli spaces of conformal field theories with $c \geq 1$ ", preprint THU-87/30.
- [5] J.J. Atick and E. Witten, Nucl. Phys. B310 (1988) 291; R. Brandenberger and C. Vafa, Nucl. Phys. B316 (1989) 391; A. Giveon, E. Rabinovici and G. Veneziano, Nucl. Phys. B322 (1989) 167; A. Shapere and F. Wilczek, Nucl. Phys. B320 (1989) 669; M. Dine, P. Huet and N. Seiberg, Nucl. Phys. B322 (1989) 301; J. Molera and B. Ovrut, Phys. Rev. D40 (1989) 1146; M. Duff, Nucl. Phys. B335 (1990) 610; E. Alvarez and M.A.R. Osorio, Phys. Rev. D40 (1989) 1150; J. Lauer, J. Mas and H.P. Nilles, Phys. Lett. B226 (1989) 251; Nucl. Phys. B351 (1991) 353; W. Lerche, D. Lüst and N.P. Warner, Phys. Lett. B231 (1989) 417.
- [6] S. Ferrara, D. Lüst, A. Shapere and S. Theisen, Phys. Lett. B225 (1989) 363.
- [7] B. Greene and M. Plesser, Nucl. Phys. B338 (1990) 15; A. Giveon and D.-J. Smit, Nucl. Phys. B349 (1991) 168; P.S. Aspinwall and C.A. Lütken, Nucl. Phys. B355 (1991) 482; M. Cvetič, J. Molera and B. Ovrut, Phys. Lett. B248 (1990) 83; P. Candelas, X. De La Ossa, P. Green and L. Parkes, Nucl. Phys. B359 (1991) 21; Phys. Lett. B258 (1991) 118.
- [8] K. Narain, Phys. Lett. 169B (1986) 41; K. Narain, M. Sarmadi and E. Witten, Nucl. Phys. B279 (1987) 369.

- [9] L. Dixon, J. Harvey, C. Vafa and E. Witten, Nucl. Phys. B261 (1985) 651; Nucl. Phys. B274 (1986) 285.
- [10] E. Cremmer, S. Ferrara, L. Girardello and A. Van Proeyen, Nucl. Phys. B212 (1983) 413.
- [11] E. Witten, Phys. Lett. 155B (1985) 151.
- [12] S. Ferrara, D. Lüst and S. Theisen, Phys. Lett. B233 (1989) 147.
- [13] M. Cvetič, A. Font, L.E. Ibáñez, D. Lüst and F. Quevedo, Nucl. Phys. B361 (1991) 194.
- [14] M. Dine, N. Seiberg, X. Wen and E. Witten, Nucl. Phys. B278 (1986) 769, Nucl. Phys. B289 (1987) 357.
- [15] A. Font, L.E. Ibáñez, D. Lüst and F. Quevedo, Phys. Lett. B245 (1990) 401.
- [16] L. Ibanez and D. Lüst, “*The Strong CP Problem and Target-space Modular Invariance in 4-D Strings*,” preprint CERN-TH.6120/91 (1991).
- [17] S. Ferrara, C. Kounnas, D. Lüst and F. Zwirner, “*Duality Invariant Partition Functions and Automorphic Superpotentials for (2,2) String Compactifications*,” preprint CERN-TH.6090/91 (1991).
- [18] D. Ray and I. Singer, Ann. Math. **98** (1973) 154.
- [19] I. Antoniadis, C. Bachas and C. Kounnas, Phys. Lett. B200 (1988) 297; L. Ibáñez, W. Lerche, D. Lüst and S. Theisen, Nucl. Phys. B352 (1991) 435.
- [20] A. Giveon and M. Porrati, Phys. Lett. B246 (1990) 54; Nucl. Phys. B355 (1991) 422.
- [21] L. Dixon, V. Kaplunovsky and J. Louis, Nucl. Phys. B355 (1991) 649.
- [22] H. Ooguri and C. Vafa, Mod. Phys. Lett. A5 (1990) 1389; “*Geometry of $N = 2$ strings*,” preprint EFI 91/05, HUTP-91/A003.
- [23] S. Ferrara, N. Magnoli, T.R. Taylor and G. Veneziano, Phys. Lett. B245 (1990) 409; H.P. Nilles and M. Olechowski, Phys. Lett. B248 (1990) 268; P. Binetruy and M.K. Gaillard, Phys. Lett. B253 (1991) 119.
- [24] J.P. Derendinger, L.E. Ibáñez and H.P. Nilles, Phys. Lett. 155B (1985) 65; M. Dine, R. Rohm, N. Seiberg and E. Witten, Phys. Lett. 156B (1985) 55.

- [25] D. Lüst and T.R. Taylor, Phys. Lett. B253 (1991) 335; B. Carlos, J. Casas and C. Munoz, “Massive Hidden Matter and Gaugino Condensation”, preprint CERN-TH.6049/91 (1991).
- [26] M.A. Shifman and A.I. Vainshtein, “On Holomorphic Dependence and Infrared Effects in Supersymmetric Gauge Theories”, preprint TPI-Minn-91/4-T (1991).
- [27] J.P. Derendinger, S. Ferrara, C. Kounnas and F. Zwirner, “On loop corrections to string effective field theories: field-dependent gauge couplings and sigma-model anomalies”, preprint CERN-TH.6004/91, LPTENS 91-4; J. Louis, “Non-harmonic gauge coupling constants in supersymmetry and superstring theory”, preprint SLAC-PUB-5527; V. Kaplunovsky and J. Louis, SLAC preprint to appear (1991).
- [28] V. Kaplunovsky, Nucl. Phys. B307 (1988) 145.
- [29] S. Ferrara, preprint UCLA/90/TEP/20 (1990).
- [30] S. Cecotti, S. Ferrara and L. Giradello, Int. J. Mod. Phys. 4 (1989) 2475, Phys. Lett. B213 (1988) 443; S. Ferrara and A. Strominger, in Strings '89 (R. Arnowitt, R. Bryan, M.J. Duff, D. Nanopoulos and C.N. Pope eds., World Scientific, 1990), p.245; L. Dixon, V. Kaplunovsky and J. Louis, Nucl. Phys. B329 (1990) 27; A. Strominger, Comm. Math. Phys. 133 (1990) 163; L. Castellani, R. D'Auria and S. Ferrara, Class. Quant. Grav. 7 (1990) 1767.
- [31] B. De Wit, P.G. Lauwers and A. Van Proeyen, Nucl. Phys. B255 (1985) 569.
- [32] S. Ferrara, “Superstring effective field theories on $(2, 2)$ vacua”, preprint CERN-TH.5873/90.
- [33] S. Ferrara, D. Lüst and S. Theisen, Phys. Lett. B242 (1990) 39.
- [34] E. Witten, Comm. Math. Phys. 118 (1988) 411; Nucl. Phys. B340 (1990) 281; R. Dijkgraaf and E. Witten, Nucl. Phys. B342 (1990) 486; C. Vafa, Mod. Phys. Lett. A6 (1991) 337.
- [35] R. Dijkgraaf, E. Verlinde and H. Verlinde, “Topological strings in $d < 1$ ”, preprint PUPT-1204 (1990).
- [36] W. Lerche, C. Vafa and N.P. Warner, Nucl. Phys. B324 (1989) 427.

MODULI SPACES AND THE GEOMETRIES OF N=2 SUSY

Pietro Fré and Paolo Soriani

SISSA - International School for Advanced Studies

Via Beirut 2, I-34100 Trieste, Italy

and

I.N.F.N. sezione di Trieste

Abstract

We review some recent results on special Kähler geometry and its uses in N=2 theories. In particular, we describe the structure of the N=2, D=4 scalar potential, an algorithm to obtain automorphic functions for orbifold moduli spaces and a new approach to the lagrangian of (2,2) theories

1. Introduction

One of the most striking results of supersymmetric quantum field theories is the deep relation between geometry, topology and supersymmetry.

In particular global and local supersymmetric 4D theories exhibit deep geometrical structures inherent to the nonlinear interaction of matter multiplets. Almost all of these couplings can be rephrased in a geometrical language as couplings of some nonlinear σ -models to gravity and gauge fields [1-7].

N=2, 4D supergravity is probably the richest theory from the geometrical point of view. In this case the constraints posed by supersymmetry on the scalar sector are sufficiently strong to select very "special" geometries and yet not so strong as to determine completely the scalar manifold $\mathcal{M}_{scalar}$. This latter is actually the direct product of a special Kähler manifold $\mathcal{S}$ with a quaternionic one $\mathcal{Q}$. The manifold $\mathcal{S}$ contains the scalars sitting in the vector multiplets and it is a Kähler manifold whose Riemann tensor satisfies the following identity

$$\mathcal{R}_{i^* j \ell^* k} = -g_{\ell^* (j} g_{k) i^*} + \frac{1}{2} e^{2G} W_{i^* \ell^* s^*} W_{t k j g} s^{*t} \quad (1.1)$$

W_{ijk} being a holomorphic 3-index symmetric tensor. On the other hand $\mathcal{Q}$ contains the scalars sitting in the hypermultiplets.

For a long time the general form of N=2 matter coupled supergravity with an arbitrary $\mathcal{S}$ and an arbitrary $\mathcal{Q}$ was missing in the literature. In particular a general formula for the N=2 scalar potential, analogous to the general formula of standard N=1 supergravity:

$$\begin{aligned} \mathcal{V}(z, \bar{z}) &= e^{J(z, \bar{z})} (\partial_i J \partial_{j^*} J g^{ij^*} - 3) \\ J(z, \bar{z}) &= G(z, \bar{z}) + \ln |\Delta(z)|^2 \end{aligned} \quad (1.2)$$

did not exist. For the sake of the next coming discussion we recall that the exponential of the Kähler potential $e^{G(z,\bar{z})}$ is interpreted as the metric of a line bundle $\mathcal{L}$ of which the superpotential $\Delta(z)$ is a holomorphic section. Hence the invariant function $J(z,\bar{z})$ appearing in the $N=1$ formula (1.2) for the scalar potential is nothing else but the norm squared $||\Delta||^2$ of the superpotential.

In a recent paper [8] the programme of supergravity matter coupling was finally completed by providing the general form of the $N=2$ supergravity lagrangian and the $N=2$ generalization of eq. (1.2). In geometrical language these formulae were obtained by investigating the gauging of isometries in special Kähler and quaternionic manifolds. In this talk we will present the result of [8], emphasising the crucial role of the symplectic structure underlying special geometry. This structure links these results on $N=2$ Supergravity with other results we are going to review on automorphic superpotentials [9] and lagrangian formulations of (2,2) superconformal theories [10].

Indeed special geometry is not only related to $N=2$ susy on the target manifold but also to $n=2$ susy on the world sheet. Actually special Kähler manifolds emerge as moduli spaces of (2,2) superconformal field theories or, alternatively, as moduli spaces of Calabi-Yau complex 3-folds. This follows from Gepner conjecture [11], supported by overwhelming evidences [12], that any $c = 3n$ ($n \in \mathbb{N}$) (2,2) theory is the exact quantum solution of an $n=2$ σ -model on a complex n -fold F_n of vanishing first Chern class: $c_1(F_n) = 0$. Such spaces have always $SU(n)$ holonomy and are named Calabi-Yau n -folds. In the $c = 9$ case, the complete moduli space of a (2,2) theory was proven to be the direct product of two special Kähler manifolds [13,14,15]

$$M_{moduli} = \mathcal{S}_{(1,1)} \otimes \mathcal{S}_{(2,1)} \quad (1.3)$$

where the complex dimensions of the two submanifolds ($\dim \mathcal{S}_{(1,1)} = h_{(1,1)}$; $\dim \mathcal{S}_{(2,1)} = h_{(2,1)}$) are the two independent Hodge numbers of the Calabi-Yau 3-fold. From the algebraic view-point $\mathcal{S}_{(1,1)}$ is the parameter space of the Kähler class deformations, associated with harmonic (1,1)-forms, while $\mathcal{S}_{(2,1)}$ is the parameter space of the complex structure deformations associated with harmonic (2,1)-forms.

From the abstract (2,2) viewpoint $\mathcal{S}_{(2,1)}$ is the parameter space of the superconformal theory deformations induced by chiral-chiral primary fields, characterized by conformal weights $h = \bar{h} = \frac{1}{2}$ and $U(1)$ charges $q = \bar{q} = 1$, while $\mathcal{S}_{(1,1)}$ is the parameter space of the deformations induced by chiral-antichiral primary fields having $h = \bar{h} = \frac{1}{2}$ and $q = -\bar{q} = 1$.

The relation between the supergravity and the Calabi-Yau view point on special geometry was provided in the past [13,14], through the observation that, on the same 3-fold one can compactify either the heterotic superstring or the type II superstring, displaying $N=1$ and $N=2$ supersymmetry, respectively. The moduli-space geometry must therefore be compatible with $N=2$ SUSY and that is the reason why it is special. Actually in [15] the relation (1.1) has been directly derived from (2,2) superconformal Ward identities. This has been the direct proof that special Kähler geometry is the geometry of the moduli space for (2,2) superconformal field-theories. New insight in this direction has recently been obtained [16,17] by the topological reinterpretation (through twisting) of $n=2$ 2D theories. Quite generally, in special Kähler manifolds,

one can find special coordinate patches z^i where the Kähler potential $G(z, \bar{z})$ is obtained from a holomorphic prepotential $f(z)$ through the formula:

$$G(z, \bar{z}) = -\log \left[2(f + \bar{f}) - (\partial_i f - \partial_{i^*} \bar{f}) (z^i - \bar{z}^{i^*}) \right] \quad (1.4)$$

In the twisted (2,2) models the holomorphic prepotential is related to the topological free-energy and W_{ijk} is related to the topological 3-point function. In ref [26] a generalization of special geometry has been considered suitable to include all topological deformations besides the superconformal ones. In this talk we restrict ourselves to proper special Kähler geometry. As already anticipated, the relation with Dolbeault cohomology of the underlying 3-fold puts into evidence a crucial symplectic $Sp(2+2n, R)$ structure (corresponding to change of bases for the homology cycles), which eventually can be taken as an intrinsic definition of special Kähler geometry. We shall review this definition in the next section. From the $N=2, D=4$ supergravity viewpoint, the symplectic $Sp(2+2n, R)$ structure is related with duality transformations that fit in the general scheme developed in [18]. The symplectic structure and the existence of a W_{ijk} satisfying eq.(1.1) are equivalent properties implying each other.

Duality transformations associated with a discrete subgroup $\Gamma \subset Sp(2+2n, Z) \subset Sp(2+2n, R)$ of the above mentioned symplectic group have become a focus of interest [19a,19b]. Indeed when the (2,2) superconformal theory is utilized to compactify a heterotic superstring with $N=1$ target supersymmetry, the group Γ acts as a target space modular group and corresponds to the action on the Calabi-Yau homology basis of global diffeomorphisms. For the same reason as in string perturbation theory, Γ must be modded out and it turns out to be an exact non perturbative symmetry of the effective low energy lagrangian. This provides a powerful tool to obtain exact non perturbative results for the moduli dependence of the superpotential $\Delta(z)$, in the corresponding $N=1$ supergravity Lagrangian. In order to maintain Γ symmetry $\Delta(z)$ has to be a Γ -automorphic function.

In [19a] a general formula has been proposed to express the automorphic superpotential $\Delta(z)$ as a sum over a Γ -homogeneous lattice* Λ_Γ :

$$\begin{aligned} \log ||\Delta(z)||^2 &= \log \left[|\Delta(z)|^2 e^{G(z, \bar{z})} \right] = \\ &= \left[- \sum_{(M_\Sigma, N^\Sigma) \in \Lambda_\Gamma} \log \frac{|M_\Sigma X^\Sigma + i N^\Sigma \partial_\Sigma F|^2}{X^\Sigma \bar{\partial}_\Sigma \bar{F} + \bar{X}^\Sigma \partial_\Sigma F} \right]_{reg} \end{aligned} \quad (1.5)$$

where $X^\Sigma(z)$ ($\Lambda = 0, 1, \dots, n$) and $i\partial_\Sigma F(X(z))$ define a holomorphic section of the $Sp(2+2n, R)$ flat bundle, whose existence corresponds to the intrinsic geometric characterization of special Kähler manifolds $\mathcal{S}$ [4e,4d,8], and where a suitable regularization of the infinite sum is understood. $F(X)$ is a homogeneous function of degree

* see also the talk given in this workshop by D. Lüst

two in the variables X^Σ . Special coordinates for the underlying Kähler manifold correspond to the choice $z^i = X^i/X^0$ ($i = 1, \dots, n$) and the prepotential $f(z)$ mentioned above is related to F through

$$f(z) = (X^0)^{-2} F(X) \quad (1.6)$$

With this identification the formula (1.4) for the Kähler potential becomes

$$G(z, \bar{z}) = -\log(X^\Sigma \bar{\partial}_\Sigma \bar{F} + \bar{X}^\Sigma \partial_\Sigma F) \quad (1.7)$$

In a recent paper [9] we have observed that in order to obtain an explicit evaluation of the automorphic superpotential, we need two informations:

- i) An explicit form for the homogeneous function $F(X)$
- ii) An explicit embedding of the homogeneous Γ lattice Λ_Γ into the homogeneous $Sp(2+2n, Z)$ lattice $\Lambda_{Sp(2+2n, Z)}$ which corresponds to an unrestricted sum over all the integers M_Σ, N^Σ

The embedding $\Lambda_\Gamma \subset \Lambda_{Sp(2+2n, Z)}$ is clearly determined once we know the embedding $\Gamma \subset Sp(2+2n, Z)$. In [9] we have considered the case where the special variety $S = \frac{S'}{\Gamma}$ is an orbifold of a special Kähler coset manifold $S' = G/H$ with respect to the action of a discrete subgroup $\Gamma \subset G$ of the isometry group. Such a situation is realized for the moduli spaces of all abelian orbifolds. In this case both problems (i) and (ii) can be solved in one stroke.

Indeed the function $F(X)$ is completely determined by the embedding of G into $Sp(2+2n, R)$ and of H into the maximal compact subgroup $U(n+1) \subset Sp(2+2n, R)$. This is just the general construction of continuous duality transformations according to [18]. Different choices of $F(X)$ correspond to different embeddings. Furthermore the embedding of the modular group Γ into $Sp(2+2n, Z)$ is obviously realized by restricting from R to Z .

In this talk we review the results of [9] for the case of the manifolds $SK(n+1)$ defined below:

$$SK(n+1) = \frac{SU(1,1)}{U(1)} \otimes \frac{SO(2,n)}{SO(2) \otimes SO(n)} \quad (1.8)$$

The homogeneous special Kähler manifolds were classified in [2,4c]: in addition to the family (1.8) one has the series

$$CP_{n-1,1} = \frac{SU(1,n)}{SU(n) \otimes U(1)} \quad (1.9)$$

corresponding to the so called minimal coupling and the four sporadic cases

$$\frac{Sp(6, R)}{U(3)}; \quad \frac{SU(3,3)}{SU(3) \otimes U(3)}; \quad \frac{SO^*(12)}{U(6)}; \quad \frac{E_{7(-25)}}{E_6 \otimes U(1)} \quad (1.10)$$

The second manifold in eq. (1.10) is the "Teichmüller" space for the deformations of the orbifold T^6/Z_3 and a similar construction for the corresponding symplectic structure and automorphic function is now in progress.

We now present the afore mentioned results in the following order. First we briefly recall the definition of special Kähler manifolds, then we illustrate the $N=2$ generalization of formula (1.2) for the scalar potential. Subsequently we discuss the symplectic embedding of $SK(n+1)$ and the associated automorphic superpotential. Finally we give a short account of a new formulation of (2,2) superconformal theories, in terms of b-c- β - γ systems that yields a world sheet lagrangian, explicitly depending on the complex structure deformation moduli. This formulation might shed new light on the structure of the topological free energy $F(X)$.

2. Definition of special Kähler manifolds

A special Kähler manifold is a Kähler manifold with additional restrictions on its Kähler structure. A manifold M_n , of complex dimension n , is Kählerian if it has a complex structure and a hermitean metric

$$ds^2 = g_{i\bar{j}}(z, \bar{z}) dz^i \otimes d\bar{z}^{\bar{j}} \quad (2.1)$$

such that the (1,1)-form

$$K = ig_{i\bar{j}}(z, \bar{z}) dz^i \wedge d\bar{z}^{\bar{j}} \quad (2.2)$$

is closed ($dK = 0$). As it is well known, K cannot be globally exact, yet it is certainly locally exact. Indeed in every coordinate patch we can find a real function $G(z, \bar{z})$ (named the Kähler potential) such that

$$g_{i\bar{j}} = \partial_i \partial_{\bar{j}} G(z, \bar{z}) \quad (2.3a)$$

$$K = dQ \quad (2.3b)$$

$$Q = -\frac{i}{2}(\partial_i G dz^i - \partial_{\bar{i}} G d\bar{z}^{\bar{i}}) \quad (2.3c)$$

Under a Kähler transformation

$$G \rightarrow G + f(z) + f(\bar{z}) \quad (2.4)$$

the 1-form Q transforms as

$$Q \rightarrow Q + d(Imf). \quad (2.5)$$

Therefore Q is a $U(1)$ connection in a line bundle $\mathcal{L}$. If the $U(1)$ line bundle is such that the first Chern class $c_1(\mathcal{L})$ coincides with the Kähler class $[K]$ then the Kähler manifold is of restricted type or a Hodge manifold. A special Kähler manifold is a Hodge manifold obeying additional restrictions that we shall presently illustrate. To this effect we point out that in addition to the $U(1)$ -holomorphic connection $\partial_i G$ one has the holomorphic Levi Civita connection and its curvature.

By definition a restricted Kähler manifold is *special* if and only if there exist a completely symmetric holomorphic 3-index section W_{ijk} of $(T^*)^3 \otimes \mathcal{L}^2$ (and its

antiholomorphic conjugate W_{i^*j,k^*}) such that eq.(1.1) is satisfied. In [4e,4d,8] it was shown that on a n -dimensional special Kähler manifold one can always introduce a $(n+1)$ - dimensional holomorphic vector bundle whose holomorphic sections we denote by X^Λ ($\Lambda = 1, \dots, n+1$) and a function $F(L)$ which is holomorphic and homogeneous of degree two in the transformed section

$$L^\Lambda(z, \bar{z}) = e^{1/2G(z, \bar{z})} X^\Lambda(z) \quad (2.6)$$

This means that $F(L) = e^{G(z, \bar{z})} F(X(z))$, so that $F(X)$ is a holomorphic section of $\mathcal{L}^2$. The geometry of the special manifold is completely determined by the sections $\{X^\Lambda\}$ and by the analytic function $F(L)$ or equivalently by $F(X)$. Define

$$F_{\Lambda_1 \dots \Lambda_n} = \frac{\partial}{\partial \Lambda_1} \frac{\partial}{\partial \Lambda_2} \dots \frac{\partial}{\partial \Lambda_n} F(L) \quad (2.7)$$

and set

$$N_{\Lambda\Sigma} = F_{\Lambda\Sigma} + F_{\Lambda\Sigma} \quad (2.8)$$

It turns out that $\Omega = \{X^\Lambda, i \frac{\partial F}{\partial X^\Lambda}\}$ can be viewed as the section of a flat holomorphic, $Sp(2n+2, R)$ -bundle $\mathcal{H}$ [4e] such that the Kähler 2-form is given by:

$$K = -\partial\bar{\partial} \log (-i < \bar{\Omega} | \Omega >) \quad (2.9)$$

where $< \bar{\Omega} | \Omega >$ denotes the compatible hermitean metric on $\mathcal{H}$:

$$< \bar{\Omega} | \Omega > = \Omega^\dagger \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \Omega = i (\bar{X}^\Lambda \partial_\Lambda F + X^\Lambda \bar{\partial}_\Lambda \bar{F}) \quad (2.10)$$

Equation (2.10) is equivalent to the following formula relating the Kähler potential G to the norm of the holomorphic section Ω in the $Sp(2n+2)$ flat bundle :

$$G = -\log \|\Omega\|^2 \equiv -\log (-i < \bar{\Omega} | \Omega >) = -\log (N_{\Lambda\Sigma} X^\Lambda \bar{X}^\Sigma) \quad (2.11)$$

3. The N=2 SUGRA Lagrangian and the scalar potential

Let us now present the N=2 Bosonic Lagrangian and the N=2 analogue of eq. (1.2), namely the general formula for the scalar potential. The result obtained in [8] is the following:

$$\begin{aligned} \mathcal{L}_{bosonic}^{N=2} = & \sqrt{-g} \left\{ \mathcal{R} + g_{ij^*}(z, \bar{z}) \nabla_\mu z^i \nabla_\nu \bar{z}^{j^*} g^{\mu\nu} \right. \\ & + h_{uv}(q) \nabla_\mu q^u \nabla_\nu q^v g^{\mu\nu} - 4 \operatorname{Re} \mathcal{N}_{\Lambda\Sigma}(z, \bar{z}) F_{\mu\nu}^\Lambda F^{\Sigma\mu\nu} \\ & \left. - \mathcal{V}(z, \bar{z}, q) \right\} - 2i \operatorname{Im} \mathcal{N}_{\Lambda\Sigma}(z, \bar{z}) F_{\mu\nu}^\Lambda F_{\rho\sigma}^\Sigma \epsilon^{\mu\nu\rho\sigma} \end{aligned} \quad (3.1)$$

where

$$\nabla_\mu z^i = \partial_\mu z^i + g A_\mu^\Lambda k_\Lambda^i(z) \quad (3.2a)$$

$$\nabla_\mu \bar{z}^{j*} = \partial_\mu \bar{z}^{j*} + g A_\mu^\Lambda k_\Lambda^{j*}(\bar{z}) \quad (3.2b)$$

$$\nabla_\mu q^u = \partial_\mu q^u + g A_\mu^\Lambda k_\Lambda^u(q) \quad (3.2c)$$

are the covariant differentials of the special scalars $z^i \in \mathcal{S}$ and of the quaternionic scalars $q^\mu \in \mathcal{Q}$. The 2-tensor

$$\mathcal{N}_{\Lambda\Sigma} = -\bar{F}_{\Lambda\Sigma} + \frac{1}{N_{\Gamma\Delta} L^\Gamma L^\Delta} N_{\Lambda\Pi} L^\Pi N_{\Sigma\Delta} L^\Delta \quad (3.3)$$

is the kinetic metric of the vectors constructed in terms of the L-sections of the $Sp(2n+2)$ -bundle. Finally the formula for the scalar potential is the following one:

$$\begin{aligned} \mathcal{V}(z, \bar{z}, q) = g^2 & \left[\left(g_{ij*} k_\Lambda^i k_\Sigma^{j*} + h_{uv} k_\Lambda^u k_\Sigma^v \right) \bar{L}^\Lambda L^\Sigma \right. \\ & \left. - \left([N^{-1}]^{\Lambda\Sigma} + 2\bar{L}^\Lambda L^\Sigma \right) \mathcal{P}_\Lambda^z \mathcal{P}_\Sigma^z \right] \end{aligned} \quad (3.4)$$

in terms of the Killing vectors $\vec{k}_\Lambda$ and of their prepotentials $\mathcal{P}_\Lambda^z$.

To explain what we mean by this we point out that the gauge group G acts as a group of isometries on the scalar manifold $\mathcal{S} \otimes \mathcal{Q}$. Let $\mathbf{G}$ be its Lie algebra. On $\mathcal{S}$ the isometries are generated by the killing vectors:

$$\vec{k}_\Lambda = k_\Lambda^i \vec{\partial}_i + k_\Lambda^{i*} \vec{\partial}_{i*} \quad (3.5)$$

whose commutation relations

$$[\vec{k}_\Lambda, \vec{k}_\Sigma] = -f_{\Lambda\Sigma}^{\Gamma} \vec{k}_\Gamma \quad (3.6)$$

define the structure constants $f_{\Lambda\Sigma}^{\Gamma}$ of G . In a general Kähler manifold $\mathcal{S}(n)$ the dimension $d = \dim \mathbf{G}$ of the isometry algebra has no particular relation with the dimension n of the manifold. In a special Kähler manifold $\mathcal{S}$ one has $d = n + 1$ and the Killing vectors are in one-to-one correspondence with the components of the holomorphic section $\{X^\Lambda(z)\}$. It is for this reason that they have been labelled with the same capital greek index.

Invariance of the Kähler 2-form under the action of $\mathbf{G}$ implies:

$$\ell_\Lambda K \equiv_\Lambda |dK + d(\lrcorner_\Lambda K) = 0 \quad (3.7)$$

where ℓ_Λ and $\lrcorner_\Lambda$ denote, respectively, the Lie derivative and the contraction along $\vec{k}_\Lambda$. Since K is closed, we conclude that $\lrcorner_\Lambda K$ is locally exact, so that in every coordinate patch we can find a real function $\mathcal{P}_\Lambda^o(z, \bar{z})$ such that:

$$\lrcorner_\Lambda K = -d\mathcal{P}_\Lambda^o \quad (3.8)$$

In coordinates this means:

$$\begin{aligned}k_{\Lambda}^i(z) &= ig^{ij*} \partial_j \mathcal{P}_{\Lambda}^{\circ}(z, \bar{z}) \\ k_{\Lambda}^{i*}(\bar{z}) &= -ig^{i*j} \partial_j \mathcal{P}_{\Lambda}^{\circ}(z, \bar{z})\end{aligned}$$

The real function $\mathcal{P}_{\Lambda}^{\circ}(z, \bar{z})$ is called the Killing vector prepotential and, in the applications of Kähler geometry to $N = 1$ supergravity, it is essentially the value taken by the auxiliary field D_{Λ} of the gauge vector multiplets.

From a geometrical point of view the prepotential $\mathcal{P}_{\Lambda}^{\circ}$ is the Hamiltonian function, providing the Poissonian realization of the Lie algebra on the Kähler manifold.

Indeed the very existence of the closed 2-form K guarantees that every Kähler space is a symplectic manifold and that we can define a Poisson bracket.

To every generator of the abstract Lie algebra $\mathbf{G}$ we have associated a function $\mathcal{P}_{\Lambda}^{\circ}$ on $\mathcal{S}$; the Poisson bracket of $\mathcal{P}_{\Lambda}^{\circ}$ with $\mathcal{P}_{\Sigma}^{\circ}$ is defined as follows:

$$\{\mathcal{P}_{\Lambda}^{\circ}, \mathcal{P}_{\Sigma}^{\circ}\} \equiv 2K(\Lambda, \Sigma) \quad (3.9)$$

where $K(\Lambda, \Sigma) \equiv K(\vec{k}_{\Lambda}, \vec{k}_{\Sigma})$ is the value of K along the pair of Killing vectors. One can prove that the following identity is true [8]:

$$\{\mathcal{P}_{\Lambda}^{\circ}, \mathcal{P}_{\Sigma}^{\circ}\} = f_{\Lambda\Sigma}^{\Gamma} \mathcal{P}_{\Gamma}^{\circ} + C_{\Lambda\Sigma} \quad (3.10)$$

where $C_{\Lambda\Sigma}$ is a constant fulfilling the cocycle condition

$$f_{\Lambda\Pi}^{\Gamma} C_{\Gamma\Sigma} + f_{\Pi\Sigma}^{\Gamma} C_{\Gamma\Lambda} + f_{\Sigma\Lambda}^{\Gamma} C_{\Gamma\Pi} = 0 \quad (3.11)$$

If the Lie Algebra $\mathbf{G}$ has a trivial second cohomology group $H^2(\mathbf{G}) = 0$ then the cocycle $C_{\Lambda\Sigma}$ is a coboundary, namely we have :

$$C_{\Lambda\Sigma} = f_{\Lambda\Sigma}^{\Gamma} C_{\Gamma} \quad (3.12)$$

where C_{Γ} are suitable constants. Hence, assuming $H^2(\mathbf{G}) = 0$ we can reabsorb C_{Γ} in the definition of $\mathcal{P}_{\Lambda}^{\circ}$:

$$\mathcal{P}_{\Lambda}^{\circ} \rightarrow \mathcal{P}_{\Lambda}^{\circ} + C_{\Lambda} \quad (3.13)$$

and we obtain the stronger equation

$$\{\mathcal{P}_{\Lambda}^{\circ}, \mathcal{P}_{\Sigma}^{\circ}\} = f_{\Lambda\Sigma}^{\Gamma} \mathcal{P}_{\Gamma}^{\circ} \quad (3.14)$$

Note that $H^2(\mathbf{G}) = 0$ is true for all semisimple Lie Algebras.

It is an easy matter to write down (3.14) in components:

$$\frac{i}{2} g_{ij*} (k_{\Lambda}^i k_{\Sigma}^{j*} - k_{\Sigma}^i k_{\Lambda}^{j*}) = -\frac{1}{2} f_{\Lambda\Sigma}^{\Gamma} \mathcal{P}_{\Gamma}^{\circ} \quad (3.15)$$

Eq. (3.14) has an analogue on quaternionic manifolds. Such an identity is crucial in proving the supersymmetry invariance of the $N = 2$ supergravity Lagrangian.

All the Killing vector properties proved so far are true on any Kähler manifold. In the case of a special Kähler manifold there is an additional condition to be imposed on $\bar{k}_\Lambda$. We must require that the sections $\{X^\Lambda\}$ of the holomorphic $(n+1)$ -bundle should transform in the adjoint representation of G :

$$\ell_\Lambda X^\Gamma = k_\Lambda^i \partial_i X^\Gamma = -f_{\Lambda\Sigma}^\Gamma X^\Sigma \quad (3.16)$$

Furthermore one has the identities:

$$\begin{aligned} L^\Lambda k_\Lambda^i &= \bar{L}^\Lambda k_\Lambda^{i*} = 0 \\ L^\Lambda \mathcal{P}_\Lambda^\circ &= \bar{L}^\Lambda \mathcal{P}_\Lambda^\circ = 0 \end{aligned} \quad (3.17)$$

Prepotentials for the Killing vectors exist also on quaternionic manifolds. Let us first summarize the definition and the properties of a quaternionic Kähler manifold $\mathcal{Q}(m)$ [20].

$\mathcal{Q}(m)$ is a $4m$ -dimensional real manifold endowed with a metric h :

$$ds^2 = h_{uv}(q) dq^u \otimes dq^v \quad (3.18)$$

and three complex structures $(J^x)_v^u$ ($x = 1, 2, 3$) that satisfy the quaternionic algebra

$$J^x J^y = -\delta^{xy} + \epsilon^{xyz} J^z \quad (3.19)$$

Defining the three 2-forms

$$\begin{aligned} \Omega^x &= \Omega_{uv}^x dq^u \wedge dq^v \\ \Omega_{uv}^x &= \lambda h_{uw} (J^x)_v^w \end{aligned} \quad (3.20)$$

where λ is some real parameter, we complete the definition of a quaternionic manifold by assuming that Ω^x are covariantly closed

$$\nabla \Omega^x \equiv d\Omega^x + \epsilon^{xyz} \omega^y \wedge \Omega^z = 0 \quad (3.21)$$

with respect to an $SU(2)$ connection ω^x such that Ω^x is its field strength:

$$d\omega^x + \frac{1}{2} \epsilon^{xyz} \omega^y \wedge \omega^z = \Omega^x \quad (3.22)$$

As a consequence of this structure, the manifold $\mathcal{Q}(m)$ has a holonomy group $\text{Hol}(\mathcal{Q}(m)) = SU(2) \otimes \mathcal{H}$ where $\mathcal{H} \subset Sp(2m)$ is some subgroup of the symplectic group in $2m$ -dimensions.

As already stated, in applications to $N = 2$ supergravity we must assume that on $\mathcal{Q}(m)$ we have an action by quaternionic isometries of the same $(n+1)$ -dimensional Lie group G that acts on the Special Kähler manifold $S(n)$. This means that on $\mathcal{Q}(m)$ we have Killing vectors

$$\bar{k}_\Lambda = k_\Lambda^u \frac{\bar{\partial}}{\partial q^u} \quad (3.23)$$

satisfying the same Lie algebra (3.6) as the corresponding Killing vectors on $S(n)$. In different words

$$\hat{k}_\Lambda = \tilde{k}_\Lambda = k_\Lambda^i \tilde{\partial}_i + k_\Lambda^{i*} \tilde{\partial}_{i^*} + k_\Lambda^u \tilde{\partial}_u \quad (3.24)$$

is a Killing vector of the block diagonal metric:

$$\hat{g} = \left(\begin{array}{c|c} g_{ij^*} & 0 \\ \hline 0 & h_{uv} \end{array} \right) \quad (3.25)$$

defined on the product manifold $S \otimes Q$.

In order to be compatible with the quaternionic structure of the manifold, the Killing vectors must leave invariant the 2-forms Ω^x up to $SU(2)$ gauge transformations. Namely:

$$\begin{aligned} \ell_\Lambda \Omega^x &= \epsilon^{xyz} \Omega^y W_\Lambda^z \\ \ell_\Lambda \omega^x &= \nabla W_\Lambda^x \end{aligned} \quad (3.26)$$

where W_Λ^x is an $SU(2)$ compensator associated with the Killing vector k_Λ^u . The compensator W_Λ^x necessarily fulfills the cocycle condition:

$$\ell_\Lambda W_\Sigma^x - \ell_\Sigma W_\Lambda^x + \epsilon^{xyz} W_\Lambda^y W_\Sigma^z = f_{\Lambda\Sigma}^\Gamma W_\Gamma^x \quad (3.27)$$

In a fully analogous way to the case of Kähler manifolds, to each quaternionic Killing vector we can associate a triplet $\mathcal{P}_\Lambda^x(q)$ of 0-form prepotentials.

Indeed on $Q(m)$ eq.(3.8) is replaced by

$$\Lambda | \Omega^x = -\nabla \mathcal{P}_\Lambda^x \equiv -(d\mathcal{P}_\Lambda^x + \epsilon^{xyz} \omega^y \mathcal{P}_\Lambda^z) \quad (3.28)$$

where ∇ denote the $SU(2)$ covariant exterior derivative.

In the quaternionic case $\mathcal{P}_\Lambda^x$ is explicitly constructed in terms of the compensator W_Λ^x . Indeed we can write

$$\mathcal{P}_\Lambda^x = H_\Lambda^x - W_\Lambda^x \quad (3.29)$$

where we have set

$$H_\Lambda^x = \Lambda | \omega^x \quad (3.30)$$

$$H_\Lambda^x = \Lambda | \omega^x \quad (3.31)$$

Again in analogy with the special Kähler case we can define a Poisson bracket also on the quaternionic prepotentials $\mathcal{P}_\Lambda^x$. We set:

$$\{\mathcal{P}_\Lambda, \mathcal{P}_\Sigma\}^x \equiv 2\Omega^x(\Lambda, \Sigma) - \frac{1}{2} \epsilon^{xyz} \mathcal{P}_\Lambda^y \mathcal{P}_\Sigma^z, \quad (3.32)$$

where $\Omega^x(\Lambda, \Sigma) \equiv \Omega^x(\vec{k}_\Lambda, \vec{k}_\Sigma)$.

To see that eq. (3.32) does indeed define an operation on the prepotentials $\mathcal{P}_\Lambda^x$ it suffices to show that eq. (3.28) can be solved for the Killing vectors in terms of the prepotentials:

$$k_\Lambda^u = \frac{1}{\lambda^2} \sum_{z=1}^3 h^{vw} (\nabla_v \mathcal{P}_\Lambda^x \Omega_{wt}^x) h^{tu} \quad (3.33)$$

Moreover the following identity is true

$$\{\mathcal{P}_\Lambda, \mathcal{P}_\Sigma\}^x = f_{\Lambda\Sigma}^\Gamma \mathcal{P}_\Gamma^x + C_{\Lambda\Sigma}^x \quad (3.34)$$

where $C_{\Lambda\Sigma}^x$ is a covariant constant

$$\nabla C_{\Lambda\Sigma}^x = 0 \quad (3.35)$$

fulfilling the cocycle condition (3.11).

This shows that $C_{\Lambda\Sigma}^x$ is a cocycle in a quaternionic valued cohomology of the Lie Algebra $\mathcal{G}$. If we assume that the second of these cohomology groups is trivial than we have

$$C_{\Lambda\Sigma}^x = f_{\Lambda\Sigma}^\Gamma C_\Gamma^x \quad (3.36)$$

and C_Γ^x can be reabsorbed in the definition of $\mathcal{P}_\Lambda^x$. In this case and only in this case we obtain the analogue of eq. (3.15):

$$\Omega_{uv}^x k_\Lambda^u k_\Sigma^v - \frac{1}{2} \epsilon^{xyz} \mathcal{P}_\Lambda^y \mathcal{P}_\Sigma^z + \frac{1}{2} f_{\Lambda\Sigma}^\Gamma \mathcal{P}_\Gamma^x = 0 \quad (3.37)$$

Eq.s (3.37) and (3.15) are the Ward identities of the $N = 2$ supersymmetry, ensuring that in the supergravity Lagrangian the variation of the bilinear Fermi terms are canceled by the variation of a suitable scalar potential. If eq. (3.37) (or (3.15)) do not hold, no scalar potential can do the work and $N = 2$ Supergravity cannot be gauged.

As one sees the scalar potential formula (3.4) is completely geometrical and it is solely written in terms of $\tilde{k}_\Lambda$, of its prepotential $\mathcal{P}_\Lambda^x$, and of the symplectic section $(L^A, \partial F)$ characterizing the special geometry

4. The symplectic embedding of $SK(n+1)$ and the new linear section

In the construction of the $N=2$ SUGRA lagrangian is essential to know the degree two, homogeneous, holomorphic function $F(X)$ and its related symplectic section. It is clear that the choice of $(X, \partial F(X))$ is not unique. Different sections are related by symplectic $Sp(2n+2, R)$ transformations [4e]. In the sequel we present our result [9], which gives a systematic way to construct the symplectic section for the case at hand (but this construction is completely general), starting from the embedding

of $SK(n+1)$ into $Sp(2(n+2), R)$. Different embeddings yield different $F(X)$'s. In the literature [4c], the F-function for the $SK(n+1)$ manifold is of the following form

$$F(X) = id_{\Lambda\Sigma\Gamma} \frac{X^\Lambda X^\Sigma X^\Gamma}{X^0} \quad (4.1)$$

where $d_{\Lambda\Sigma\Gamma}$ are constant coefficients. Explicitly one has:

$$d_{\Lambda\Sigma\Gamma} = \begin{cases} d_{1ij} = -\frac{1}{2}\eta_{ij} \\ 0 & \text{otherwise} \end{cases} \quad (4.2)$$

where the range of the index Λ is as follows

$$\Lambda = 0, I; \quad I = 1, i; \quad i = 2, \dots, n+1$$

and

$$\eta_{ij} = \text{diag}\{+, -, -, \dots, -\} \quad (4.3)$$

From equation (4.3) one observes that F is invariant under an $SO(n-1, 1)$ group and not under an $SO(n)$ group. This means that the holomorphic symplectic section $(X, i\partial F)$ does not transform linearly under $SO(n)$. Comparing with the results of [8] one concludes that if $SK(n+1)$ is utilized as the vector multiplet manifold $\mathcal{VM}$ for an $N=2$ supergravity model, one can not gauge an n -dimensional semisimple group. Apparently one always needs at least two spectator multiplets: one is sitting in $SU(1,1)/U(1)$ and the other in $SO(2, n)/SO(n) \otimes SO(2)$. Indeed the condition to gauge an $n+1$ -dimensional group G of isometries of $\mathcal{VM}$ is given by (3.16)

On the other hand, from string theory, one knows that there are $N=2$ models based on $\mathcal{VM} = SK(n+1)$ where an n -dimensional semisimple group is gauged, the only spectator being the vector multiplet that contains the dilaton and the axion ($SU(1,1)/U(1)$ factor). For instance in the free fermion construction of superstring vacua, one obtains $N=2$ models by means of a Z_2 projection on $N=4$ models, such that the massless vector multiplets are all untwisted states. This implies that the corresponding $N=2$ manifold $\mathcal{VM}$ is obtained by disintegration of the known $N=4$ scalar manifold $M^{N=4} = SO(6, \dim G^{N=4})/SO(6) \otimes SO(\dim G^{N=4})$ and one finds $\mathcal{VM} = SK(\dim G^{N=2} + 1)$ [21,22].

If $G^{N=2}$ is semisimple we are in trouble. This trouble occurs in hundreds of examples: one example is described in detail in the appendix B of the chapter VI in [18b], where $n = 418 = 3+3+3+28+133+248$ and $G^{N=2} = SU(2)^3 \otimes SO(8) \otimes E_7 \otimes E'_8$.

What we have just now described is a paradox. Its obvious solution is that there must be a different description of the special structure of $SK(n+1)$ in terms of a different section $(X, i\partial F)$ such that the $SO(n)$ transformations are linearly realized.

This section can be derived from the symplectic embedding of $SK(n+1)$, which is strongly related to the Gaillard–Zumino construction of lagrangians possessing duality rotations on the vector fields [18]. Let us recall some crucial points of this construction.

Consider lagrangian densities of the following form:

$$\mathcal{L} = -4\text{Re}(\mathcal{N}_{\Lambda\Sigma}(\phi)F_{ab}^\Lambda F^{\Sigma ab}) - 2i\text{Im}(\mathcal{N}_{\Lambda\Sigma}(\phi)F_{ab}^\Lambda F_{cd}^\Sigma \epsilon^{abcd}) + g_{\alpha\beta}\partial_\mu\phi^\alpha\partial_\mu\phi^\beta \quad (4.4)$$

They describe a system of $n+1$ -abelian vector fields A_a^Λ ($\Lambda = 0, \dots, n$) and $2m$ scalar fields ϕ_α ($\alpha = 1, \dots, 2m$). They can be obtained from the general formula of the $N = 2$ Lagrangian (3.1) by setting to zero the gauge coupling(s), by defining $z^i = \phi^i + i\phi^{m+i}$ ($i = 1, \dots, m$), and by deleting the gravitational and hypermultiplet sectors.

If we assume that the scalar manifold is a coset $\frac{G}{H}$ and that $g_{\alpha\beta}$ is its own invariant metric, then the scalar sector of the lagrangian is invariant under G -isometries. However in the case of (4.4) there is something more than the G -isometries. According to the results of [18] the isometry group G of the scalar σ -model acts on the $n+1$ -vector fields as a group of duality transformations. This means the following:

i) There is an inclusion mapping

$$G \rightarrow Usp(n+1, n+1) = U(n+1, n+1) \cap Sp(2n+2, C) \quad (4.5)$$

$$H \rightarrow U(n+1) \quad (4.6)$$

which to each element $g \in G$ associates a $(2n+2) \times (2n+2)$ complex matrix $S(g)$ with the following block structure :

$$\begin{pmatrix} \psi_0(g) & \psi_1^*(g) \\ \psi_1(g) & \psi_0^*(g) \end{pmatrix} \quad (4.7)$$

where the $(n+1) \times (n+1)$ matrices ψ_0, ψ_1 fulfill the conditions:

$$\begin{aligned} \psi_0^\dagger \psi_0 - \psi_1^\dagger \psi_1 &= 1 \\ \psi_0^\dagger \psi_1^* - \psi_1^\dagger \psi_0^* &= 0 \end{aligned} \quad (4.8)$$

ii) The equation of motion and Bianchi identities of the $n+1$ -vector field A_a^Λ can be written as a system of $2n+2$ -equations of the following type:

$$\begin{aligned} \partial^a F_{ab}^+ &= 0 \\ \frac{1}{2} \partial_a \frac{\partial \mathcal{L}}{\partial F_{ab}^+} &= 0 \end{aligned} \quad (4.9)$$

where F_{ab}^+ denotes the self dual part of F_{ab} . Analogous equations hold for the antiself dual part. Equations (4.9) are left invariant by the transformations of G defined through the action of the matrix $S(g)$.

Given a parametrization $L(\phi)$ of the scalar coset manifold $\frac{G}{H}$, we can construct immediately $S(L(\phi))$. The lagrangian, which yields equations of motion with the invariance properties (4.9) has, in terms of $\psi_0(L(\phi)) = \psi_0$ and $\psi_1(L(\phi)) = \psi_1$, a universal structure. Indeed one must have:

$$\mathcal{L} = -4(F_{\Lambda ab}^+ F_{\Sigma ab}^+ \mathcal{N}^{\Lambda\Sigma} + c.c.) \quad (4.10)$$

where $\mathcal{N}^{\Lambda\Sigma} = \text{Re}(\mathcal{N}^{\Lambda\Sigma}) + i\text{Im}(\mathcal{N}^{\Lambda\Sigma})$ and

$$-4\mathcal{N} = [\psi_0^\dagger + \psi_1^\dagger]^{-1} [\psi_0^\dagger - \psi_1^\dagger] \quad (4.11)$$

In all this construction a crucial role is played by the embedding of $\frac{G}{H}$ into $Usp(n+1, n+1)$, which defines the right structure of duality rotations in connection with G -isometries.

In the same way we are going to show that the right duality transformations on the section $(X, i\partial F)$ for the coset $SK(n+1)$ are induced by the G -isometries in the appropriate $Sp(2(n+2), R)$ embedding.

We start by recalling some basic definitions.

A general matrix of $Sp(2(n+2), R)$ can be written in the following block structure (each block is a $(n+2) \times (n+2)$ matrix):

$$\begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{pmatrix} \quad (4.12)$$

and it is defined by the condition

$$\begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{pmatrix}^T \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (4.13)$$

which implies $\mathcal{A}^T \mathcal{C} - \mathcal{C}^T \mathcal{A} = 0$; $\mathcal{B}^T \mathcal{D} - \mathcal{D}^T \mathcal{B} = 0$; $\mathcal{A}^T \mathcal{D} - \mathcal{C}^T \mathcal{B} = 1$

The manifold $\frac{SO(2, n)}{SO(n) \otimes SO(2)}$ can be described by the following equation in $CP(1, n)$:

$$\begin{aligned} \eta_{\Lambda\Sigma} Y^\Lambda Y^\Sigma &= 0 \\ \eta_{\Lambda\Sigma} Y^\Lambda \bar{Y}^\Sigma &= \frac{1}{2} \end{aligned} \quad (4.14)$$

where we set $\Lambda = 0, 1, \alpha$; $\alpha = 2, \dots, n+1$; $\eta_{\Lambda\Sigma} = (+, +, -, \dots, -)$. The constraints (4.14) are easily solved by choosing the so called Calabi-Visentini parametrization, namely:

$$\begin{aligned} Y^0 &= \frac{1}{2}(1 + y^\alpha y^\alpha)/J_1^{\frac{1}{2}} \\ Y^1 &= \frac{i}{2}(1 - y^\alpha y^\alpha)/J_1^{\frac{1}{2}} \\ Y^\alpha &= y^\alpha/J_1^{\frac{1}{2}} \\ J_1 &= (1 - 2y^\alpha \bar{y}^\alpha + y^\alpha y^\alpha \bar{y}^\beta \bar{y}^\beta) \end{aligned} \quad (4.15)$$

Finally $SU(1, 1)/U(1)$ is parametrized by choosing two complex numbers ϕ_0, ϕ_1 such that

$$|\phi_0|^2 - |\phi_1|^2 = \frac{1}{2} \quad (4.16)$$

In particular equation (4.16) is automatically satisfied if we choose

$$\begin{aligned} \phi_0 &= -\frac{D+i}{2J_2^{\frac{1}{2}}} \\ \phi_1 &= -\frac{D-i}{2J_2^{\frac{1}{2}}} \\ J_2 &= -i(D - \bar{D}) \end{aligned} \quad (4.17)$$

where $D = iS$ is the complex field containing the dilaton and the axion fields.

As already stated, in the old parametrization, the special structure of the manifold $SK(n+1)$ was realized by an F-function of type (4.1) with $d_{\Lambda\Sigma\Gamma}$ as in (4.2), where $X^0 = 1$, $X^1 = D$, $X^i = z^i$ (i runs from $2 \cdots n+1$ with lorentzian metric $(+, -, \dots, -)$). The special coordinates were defined as follows

$$\begin{aligned} z^1 &= \frac{X^1}{X^0} = D \\ z^i &= \frac{X^i}{X^0} \end{aligned} \quad (4.18)$$

and the Kähler potential was given by:

$$G(D, \bar{D}, z, \bar{z}) = -\log\left[\frac{i}{2}(D - \bar{D})(z^i - \bar{z}^i)\eta_{ij}(z^j - \bar{z}^j)\right] \quad (4.19)$$

The z -coordinates are an alternative parametrization of the manifold $\frac{SO(2,n)}{SO(n) \otimes SO(2)}$ and they are related to the Calabi-Visentini frame through an appropriate holomorphic coordinate transformation [23]. Obviously the special coordinates (4.18) are not unique; they depend on the choice of the section we consider. This means that there is a different choice of special coordinates wick as well as the z^i 's reveal the special structure of $SK(n+1)$. Let us construct the new coordinates (or equivalently the new $Sp(2(n+2), R)$ section). We consider a general $SO(2, n)$ matrix :

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad (4.20)$$

where A is a 2×2 matrix; B a $2 \times n$; C a $n \times 2$ and D a $n \times n$ matrix. A, B, C, D satisfy the following conditions:

$$\begin{aligned} A^T A - C^T C &= \mathbf{1}_{2 \times 2} \\ A^T B - B^T C &= 0 \\ A^T B - C^T D &= 0 \\ B^T B - D^T D &= -\mathbf{1}_{n \times n} \end{aligned} \quad (4.21)$$

As can be easily checked the $2(n+2) \times 2(n+2)$ matrix

$$A_1 = \begin{pmatrix} A & 0_{2 \times n} & 0_{2 \times 2} & -B \\ 0_{n \times 2} & D & -C & 0_{n \times n} \\ 0_{2 \times 2} & -B & A & 0_{2 \times n} \\ -C & 0_{n \times n} & 0_{n \times 2} & D \end{pmatrix} \quad (4.22)$$

is in $Sp(2(n+2), R)$, and it is a good candidate for the embedding we search. Moreover let us take the $SL(2, R) \sim SU(1, 1)$ matrix:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (4.23)$$

where $ad - bc = 1$. $SL(2, R)$ can be easily embedded into $Sp(2(n+2), R)$ by writing (in the same block structure notation):

$$A_2 = \begin{pmatrix} a\mathbf{1}_2 & 0 & b\mathbf{1}_n & 0 \\ 0 & d\mathbf{1}_n & 0 & c\mathbf{1}_n \\ c\mathbf{1}_2 & 0 & d\mathbf{1}_2 & 0 \\ 0 & b\mathbf{1}_n & 0 & a\mathbf{1}_n \end{pmatrix} \quad (4.24)$$

with $\mathbf{1}_2 = \mathbf{1}_{2 \times 2}$, $\mathbf{1}_n = \mathbf{1}_{n \times n}$. One can check that the two $Sp(2(n+2), R)$ matrices A_1, A_2 commute, and that they close the $SO(2, n) \otimes SL(2, R)$ algebra. Let us introduce the shorthand notation $\tilde{Y} \equiv (Y^0 J_1^{\frac{1}{2}}, Y^1 J_1^{\frac{1}{2}})$, $\hat{Y} = \{Y^\alpha J_1^{\frac{1}{2}}\}$ for the $\frac{SO(2, n)}{SO(n) \otimes SO(2)}$ variables (suitably rescaled). Under the action of an $SO(2, n)$ isometry $(\bar{Y}, \bar{Y})$ transforms as a vector. Analogously $(\phi_0 J_2^{\frac{1}{2}}, \phi_1 J_2^{\frac{1}{2}})$ transform as a vector under the action of an $SU(1, 1)$ matrix. Equivalently $(\eta_0 = (\phi_0 + \phi_1) J_2^{\frac{1}{2}}, \eta_1 = i(\phi_0 - \phi_1) J_2^{\frac{1}{2}})$ transform under the action of the $SL(2, R)$ matrix (4.23). If we impose that the section $(X, i\partial F)$ transforms as a vector under the particular $Sp(2(n+2), R)$ transformation given by $A_1 A_2$, we find the following relations:

$$\begin{aligned} \tilde{X} &= \tilde{Y} \eta_0 \\ \hat{X} &= -\hat{Y} \eta_1 \\ i\tilde{\partial} F &= \tilde{Y} \eta_1 \\ i\hat{\partial} F &= -\hat{Y} \eta_0 \end{aligned} \quad (4.25)$$

The relations (4.25) are actually a set of differential equations for F . To see this it suffices to utilize the constraints (4.14) in eq. (4.25). The solution of this system of equations is:

$$F(X) = i\sqrt{(X^0 X^0 + X^1 X^1) X^\alpha X^\alpha} \quad (4.26)$$

It is an easy algebraic calculation to verify that the function $F(X)$ in (4.26) gives the right Kähler potential.

It is now a standard matter to introduce the new special coordinates using (4.26) and to write the new holomorphic prepotential:

$$f(\pi) = i\sqrt{(1 + \pi^1 \pi^1)(\pi^\alpha \pi^\alpha)} \quad (4.27)$$

In terms of the variables π the Kähler potential is:

$$\begin{aligned} G(\pi, \bar{\pi}) &= -\log\{i[\sqrt{(1 + \pi^1 \pi^1)(\pi^\alpha \pi^\alpha)} - c.c.] - i[\pi^1 \sqrt{\frac{\pi^\alpha \pi^\alpha}{1 + \pi^1 \pi^1}} + c.c.][\pi^1 - \bar{\pi}^1] \\ &\quad + i\pi^\alpha \bar{\pi}^\alpha [\sqrt{\frac{1 + \pi^1 \pi^1}{\pi^\alpha \pi^\alpha}} - c.c.]\} \end{aligned} \quad (4.28)$$

5. Automorphic function for $SK(n+1)$

The construction of the previous section can be utilized also to define the automorphic function (1.5) for the manifold $SK(n+1)$. In this case the modular group is well known [19], and it is given by $SO(2, n, Z) \otimes PSL(2, Z)$. The only so far undefined point is how to restrict the sum over the integers M_Λ, N^Λ , which define an orbit of $Sp(2(n+2), Z)$, to unrestricted integers, defining an orbit of the true modular group $SO(2, n, Z) \otimes PSL(2, Z) \subset Sp(2(n+2), Z)$.

We consider, to perform this restriction, the conjugate transformation of (M_Λ, N^Λ) under $Sp(2(n+2), Z)$:

$$\begin{pmatrix} M' \\ N' \end{pmatrix} = \begin{pmatrix} \mathcal{D} & -\mathcal{C} \\ -\mathcal{B} & \mathcal{A} \end{pmatrix} \begin{pmatrix} M \\ N \end{pmatrix} \quad (5.1)$$

in such a way the expression (1.5) is invariant under symplectic transformations. If we denote by m^0, m^1, m^α the integers transforming as a vector under the (conjugate) $SO(2, n, Z)$ matrix; s_0, s_1 those transforming under $SL(2, Z)$ and if we use the explicit form of the matrices $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$, we get (with $\tilde{m} = (m_0, m_1)$, $\hat{m} = \{m_\alpha\}$):

$$\begin{aligned} \tilde{M} &= \tilde{m} s_0 \\ \hat{M} &= -\hat{m} s_1 \\ \tilde{N} &= \tilde{m} s_1 \\ \hat{N} &= -\hat{m} s_0 \end{aligned} \quad (5.2)$$

Equation (5.2) gives the explicit dependence of the capital integers M, N in terms of the lower case m, s , and the rule to restrict the sum in (1.5) to the appropriate $SK(n+1)$ modular group orbit. In this way one can make explicit the general formula (1.5) for the automorphic superpotential Δ . In terms of the symplectic section $\Omega = (X, i\partial F)$ eq. (1.5) can be reinterpreted as follows:

$$\log \|\Delta\|^2 = \log |\Delta|^2 e^G = - \sum_{L \in \Lambda_\Gamma} \log i \frac{|\langle L | \Omega \rangle|^2}{\langle \tilde{\Omega} | \Omega \rangle} \quad (5.3)$$

where $L = (N, -M)$ is a vector in Λ_Γ . In our case we have $L = (N, -M)$ given by eq.(5.2), and the summation on Λ_Γ corresponds to a summation over unrestricted integers (s_0, s_1) (spanning an orbit of $PSL(2, Z)$) and over integers (m_0, m_1, m_α) such that:

$$m_0^2 + m_1^2 - m_\alpha m_\alpha = 0 \quad (5.4)$$

Eq (5.4) clearly defines an orbit of $SO(2, n, Z)$ and it is immediately recognizable as the level matching condition, equating the the left and the right masses in the $\Gamma_{2,n}$ Narain Lattice.

What is still missing to make (5.3) well defined is the specification of a regularization procedure. In general this can be done via a ζ -function regularization

scheme:

$$\log |\Delta|^2 e^G = - \lim_{s \rightarrow 0} \frac{d}{ds} \zeta(s)$$

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty dt t^{s-1} \sum_{L \in \Lambda_r} \exp -it \frac{|\langle L|\Omega \rangle|^2}{\langle \Omega|\Omega \rangle} \quad (5.5)$$

This procedure always provides a correct definition for the infinite sum appearing in the right-hand side of equation (5.3). What is not obvious, in the general case, is how to extract from (5.5) the squared modulus of a holomorphic function $\Delta\Delta^*$ modulo the holomorphic anomaly $-i < \bar{\Omega}|\Omega \rangle$. In the case of $n = 2$, however, corresponding to the moduli space of certain orbifolds, we are able to work out the explicit form of Δ obtaining the product of the square of three Dedekind functions [19a]. (see also the talk given by D. Lüst)

6. Moduli dependent lagrangians for (2,2) theories

We turn now to the description of a new approach [10] to (2,2) superconformal theories that might prove useful to study their moduli space also in the topological version. We focus on $c=3n$ theories corresponding to Calabi-Yau n -folds.

Let $\Psi_\alpha\left(\frac{1}{2}, \frac{1}{2}\right)(z, \bar{z})$ ($\alpha = 1, \dots, h_{(2,1)}$) be the chiral-chiral primary fields mentioned at the beginning. Evaluating the operator product expansion with $N = 2$ supercurrents:

$$G^-(z) \tilde{G}^-(\bar{z}) \Psi_\alpha\left(\frac{1}{2}, \frac{1}{2}\right)(w, \bar{w}) = \frac{1}{|z-w|^2} \Phi_\alpha\left(\frac{1}{2}, \frac{1}{2}\right)(w, \bar{w}) + \text{reg} \quad (6.1)$$

we obtain a set of $h_{(2,1)}$ fields of conformal weights $h = \bar{h} = 1$ that can be used to deform the Lagrangian of the (2,2)-theory:

$$\mathcal{L}^{(2,2)}(z, \bar{z}) \rightarrow \mathcal{L}^{(2,2)}(z, \bar{z}) + \delta M^\alpha \Phi_\alpha \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} (z, \bar{z}) = \mathcal{L}^{(2,2)}(z, \bar{z}) \quad (6.2)$$

The parameters δM^α are the differentials of the (2,1)-moduli and the new Lagrangian $\mathcal{L}^{(2,2)}$ defines a new (2,2)-theory with the same $h_{(2,1)}$ and $h_{(1,1)}$ numbers and the same central charge.

Iterating the procedure one can reconstruct (at least in principle) a Lagrangian $\mathcal{L}^{(2,2)}(z, \bar{z}, M^\alpha)$ which explicitly depends on a set of $h_{(2,1)}$ parameters M^α and, for all values of these parameters, defines a (2,2)-theory with fixed Hodge numbers and fixed central charge.

In [10] it has been shown that there is at least one deformation class of smooth algebraic Calabi-Yau n -folds, for which the Lagrangian $\mathcal{L}^{(2,2)}(z, \bar{z}, M)$ can be explicitly written in terms of all the algebraic moduli. Furthermore the generators of the (2,2)-algebra can also be written in an explicitly moduli-dependent way and fulfill the $N = 2$ OPE's at any value of the moduli parameters.

The basic idea is that of utilizing a new set of variables, different from those of the σ -model, that capture the topological and analytic properties of the target-manifold bypassing all its metric properties. This is the same aim pursued by the

Landau-Ginzburg formulations [24] where the analytic and topological properties are encoded in the superpotential, while the metric properties, encoded in the kinetic terms, are disregarded. The difference between the Landau-Ginzburg approach and this formulation is that the former is an ordinary field theory, becoming superconformal only at some critical point, while the latter is a (2,2)-theory never moving off-criticality. In the Landau-Ginzburg Lagrangian the basic variables are scalar superfields representing the coordinates of the ambient space, in which the Calabi-Yau n -fold is immersed as a complete intersection of polynomial constraints. Classically these fields have canonical dimension $\frac{(D-2)}{2} = 0$, but quantum mechanically they acquire anomalous dimensions related with the structure of the superpotential. In our Lagrangian the basic variables are an appropriate collection of the β - γ -b-c systems whose conformal weights encode the appropriate anomalous dimensions related with the defining polynomial.

The argument leading to a formulation in terms of β - γ -b-c fields is the following. On one hand we know from Gepner's work that an interesting class of $c = 3n$ (2,2)-theories can be built as tensor product of $N = 2$ minimal models. On the other hand, it is a result shown in [10], that $n = 2$ minimal models of central charge $c = \frac{3k}{k+2}$ can be realized in terms of the supersymmetric β - γ -b-c system [25] where the conformal weights of the four fields are as follows: $h[\beta] = \lambda = \frac{1}{2k+4}$; $h[\gamma] = 1 - \lambda = \frac{2k+3}{2k+4}$; $h[b] = \lambda + \frac{1}{2} = \frac{k+3}{2k+4}$; $h[c] = \frac{1}{2} - \lambda = \frac{k+1}{2k+4}$.

Hence, for the (2,2)-theories obtained from the Gepner's tensor product construction, a Lagrangian written in terms of β - γ -b-c fields does always exist. The next question is whether $h = \tilde{h} = \frac{1}{2}$ chiral-chiral primary fields can be written solely in terms of β - γ -b-c fields. If this condition is verified, then the deformed Lagrangian is still constructed out of the same set of fields and the deformation procedure can be successfully iterated. In this way one obtains the Lagrangian $\mathcal{L}^{(2,2)}(z, \bar{z}, M)$ as a power series in the moduli parameters M^α (Noether coupling method). Surprisingly it turns out that the power series stops at the first iteration, so that the lagrangian is linear in M^α .

Consider the tensor product of N discrete series models of level k_i ($i = 1, \dots, N$) and to each of these minimal models associate a complex coordinate Z_i . Define the least integers r_i ($i = 1, \dots, N$) and d such that $r_i = \frac{d}{k_i+2}$. The (2,2)-theory obtained from the above tensor product corresponds to the $(N-2)$ -fold defined by the following degree d equation

$$WF(d, r_1, \dots, r_N): Z_1^{k_1+2} + Z_2^{k_2+2} + \dots + Z_N^{k_N+2} = 0 \quad (6.3)$$

in the weighted projective space $WCP_{N-1}(r_1, \dots, r_N)$. We name (6.3) the weighted Fermat's curve.

Note that a minimal model with $k = 0$ contributes zero to the central charge and corresponds to the identity representation of the superconformal algebra. Hence $k = 0$ factors are irrelevant in the superconformal language. However they cannot be disregarded while writing the corresponding algebraic equation (6.3). If $c = 3n$ the total number of coordinates Z_i has to be $n + 2$ in order to obtain an n -fold by means of a polynomial constraint in an $n + 1$ -weighted projective space. In a tensor

product with $N < n + 2$ the missing coordinates are associates with $k = 0$ models and have weight $\frac{d}{2}$, d being defined above. Note also that $k = 0$ factors correspond to a β - γ - b - c system with conformal weights $\frac{1}{4}, \frac{3}{4}, \frac{3}{4}, \frac{1}{4}$ (in the given order). Hence, also in the Lagrangian we are proposing $k = 0$ factors cannot be disregarded.

Clearly the weighted Fermat's curve (6.3) is a particular case of the deformation class $WCP_{N-1}(d; r_1, \dots, r_N)$, namely the class of degree d homogeneous equations in $WCP_{N-1}(d; r_1, \dots, r_N)$.

It is immediate to see the correspondence between the algebraic deformations of the defining polynomial (6.3) and the chiral-chiral primary fields that are expressible in terms of the $\beta_i - \gamma_i - b_i - c_i$ variables. It suffices to establish the following correspondence between the coordinates of the weighted projective space and the β_i -fields:

$$Z_i = \beta_i(z)\tilde{\beta}_i(\bar{z}) \quad (6.4)$$

Given a polynomial $\Pi^{(k_i+2)}(Z_i)$ of degree $k_i + 2$ in the Z_i coordinate (homogeneous of degree d in $WCP_{N-1}(r_1, \dots, r_N)$), if we perform the formal substitution (6.4), we obtain the operator $\Pi^{(k_i+2)}(\beta_i, \tilde{\beta}_i)$, that is chiral-chiral with $h = \tilde{h} = \frac{1}{2}$, $q = \tilde{q} = 1$.

Consider now a tensor product model $(k_1, k_2, \dots, k_N)$ corresponding to the weighted Fermat's curve (6.3). A world-sheet Lagrangian for this (2,2)-theory is given (in the superconformal gauge) by:

$$\begin{aligned} \mathcal{L}_0^{(2,2)}(z, \bar{z}) = \sum_{i=1}^N \frac{1}{2k_i + 4} & \left[-(\beta_i \bar{\partial} \gamma_i + \tilde{\beta}_i \partial \tilde{\gamma}_i) + (2k_i + 3)(\gamma_i \bar{\partial} \beta_i + \tilde{\gamma}_i \partial \tilde{\beta}_i) \right. \\ & \left. - (k_i + 3)(b_i \bar{\partial} c_i + \tilde{b}_i \partial \tilde{c}_i) - (k_i + 1)(c_i \bar{\partial} b_i + \tilde{c}_i \partial \tilde{b}_i) \right] \end{aligned} \quad (6.5)$$

The Lagrangian (6.5) can be shown to be invariant under a (2,2) superconformal algebra of transformations.

Consider now an arbitrary weighted homogenous polynomial $\Pi(d, r_1, \dots, r_N)$ of degree $(d, r_1, \dots, r_N)$. This polynomial can always be rewritten as: $\Pi(d, r_1, \dots, r_N) = WF(d, r_1, \dots, r_N) + M^\alpha P_\alpha(d, r_1, \dots, r_N)$, where WF is the Fermat's polynomial and where P_α is a basis of monomials for the non-trivial polynomial deformations, whose number is the number of algebraic moduli.

If we perform the substitution (6.4) we obtain that each of the operators $\Psi_\alpha \left(\begin{smallmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 1 \end{smallmatrix} \right) (z, \bar{z}) = P_\alpha(\beta_i)P_\alpha(\tilde{\beta}_i)$ is a chiral-chiral primary field with the correct weights and charges.

Inserting these Ψ_α operators into (6.1) we obtain the expression for the deformations $\Phi_\alpha \left(\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix} \right) (z, \bar{z})$ of the Lagrangian (6.5)

$$\Phi_\alpha \left(\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix} \right) (z, \bar{z}) = b_i \tilde{b}_j P_i^\alpha(\beta) \tilde{P}_j^\alpha(\tilde{\beta}) \quad (6.6)$$

where we have introduced the notation $P_{i_1 \dots i_m}^\alpha(\beta) = \frac{\partial^m}{\partial \beta_{i_1} \dots \partial \beta_{i_m}} P^\alpha(\beta)$. We conclude that for a polynomial Π , infinitesimally close to WF, the corresponding $\mathcal{L}_{(n+2)}^{(2,2)}(z, \bar{z}, M)$ lagrangian is given by:

$$\mathcal{L}_{(n+2)}^{(2,2)}(z, \bar{z}, M) = \mathcal{L}_{F_n}^{(2,2)}(z, \bar{z}) + \sum_{\alpha} M^{\alpha} b_i \bar{b}_j P_i^{\alpha} \bar{P}_j^{\alpha} \quad (6.7)$$

where $\mathcal{L}_{(F_n)}^{(2,2)}$ is defined by eq (6.5).

The lagrangian (6.7) is already exactly invariant under a full fledged (2,2) algebra of superconformal transformations. Indeed one can verify that (6.7) is invariant (up to total divergences) against the following holomorphic supersymmetry transformations ($\bar{\delta}\epsilon^{\pm} = 0$):

$$\begin{aligned} \delta\beta_i &= 2\epsilon^{-} b_i \\ \delta b_i &= \tfrac{1}{2}\epsilon^{+} \partial\beta_i + \frac{1}{2k_i + 4} \partial\epsilon^{+} \beta_i \\ \delta\gamma_i &= \tfrac{1}{2}\epsilon^{+} \partial c_i + \frac{k_i + 1}{2k_i + 4} \partial\epsilon^{+} \beta_i \\ \delta c_i &= 2\epsilon^{-} \gamma_i \\ \delta\tilde{\beta}_i &= 0 \\ \delta\tilde{b}_i &= 0 \\ \delta\tilde{c}_i &= -\tfrac{1}{2}\epsilon^{+} \sum_{\alpha} M^{\alpha} P^{\alpha} \bar{P}_i^{\alpha} \\ \delta\tilde{\gamma}_i &= \tfrac{1}{2}\epsilon^{+} \sum_{\alpha} M^{\alpha} P^{\alpha} \bar{P}_{ij}^{\alpha} b_j \end{aligned} \quad (6.8)$$

The action (6.7) is also invariant under the corresponding anti-holomorphic transformation.

The distinctive property of the above lagrangian is that the interacting fields obey the same OPE's as the non interacting ones. Furthermore the equation of motion are such that $\bar{\delta}\beta_i = \bar{\delta}b_i = 0$ as in the free theory, while $\bar{\delta}\gamma_i$ and $\bar{\delta}c_i$ are algebraic functions of the b, β fields. This allows to express the general solution of the interacting equations in terms of functionals of free fields. This fact raises hopes for a direct algebraic calculation of the topological free energy $F(M^{\alpha})$.

References

- [1] E. Cremmer, B. Julia, Nucl. Phys. B159 (1979) 141
- [2] E. Cremmer, S. Ferrara, L. Girardello, A. Van Proeyen, Nucl. Phys. B212 (1983) 413
- [3] J. Bagger and E. Witten, Nucl. Phys. B211 (1983) 302 and Phys. Lett. 115B (1982) 202
- [4a] B. de Wit and A. Van Proeyen, Nucl. Phys. B245 (1984) 89
- [4b] B. de Wit, P.G. Lauwers and A. Van Proeyen, Nucl. Phys. B255 (1985) 569

- [4c] E. Cremmer, C. Kounnas, A. Van Proeyen, J.P. Derendinger, S. Ferrara, B. de Wit and L. Girardello, Nucl. Phys. B250 (1985) 385; E. Cremmer, A. Van Proeyen, Class. and Quantum Grav. 2 (1985) 445
- [4d] L. Castellani, R. D'Auria and S. Ferrara, Phys. Lett. B241 (1990) 57 and Class. and Quantum. Grav. 1 (1990) 1767
- [4e] A. Strominger, Comm. Math. Phys. 133 (1990) 163
- [5] B. de Wit, H. Nicolai, Nucl. Phys. B208 (1982) 323
- [6] L. Castellani, A. Ceresole, R. D'Auria, S. Ferrara, P. Fre', E. Maina, Nucl. Phys. B286 (1986) 317
- [7] E. Bergshoeff, I. G. Koh and E. Sezgin, Phys. Lett. 155B (1985) 7, M. de Roo and P. Wagemans, Nucl. Phys. B262 (1985) 644
- [8] R. D'Auria, S. Ferrara, P. Fré - " Special and quaternionic isometries: general couplings in N=2 supergravity and the scalar potential " - Prepr. CERN-TH 5954/90, DFPD-90-TH-35, SISSA 192/90/EP
- [9] P. Fré, P. Soriani - " Symplectic embeddings, special Kähler geometry and automorphic functions " - Prepr. SISSA 90/91/EP
- [10] P. Fré, F. Gliozzi, M. Rego Monteiro, Annamaria Piras - " A moduli-dependent lagrangian for (2,2) theories on Calabi- Yau n-folds " - Prepr. SISSA 177/90/EP
- [11] D. Gepner, Nucl. Phys. B296 (1988) 757; Phys. Lett. B199 (1987) 380; Trieste lectures at Superstring school 1989
- [12] B.R. Greene, C. Vafa and N.P. Warner, Nucl. Phys. B324 (1989) 371
- [13] S. Cecotti, S. Ferrara, L. Girardello, Int. Jour Mod. Phys - 4 (1989) 2475 , Phys. Lett. - B213 (1988) 443
- [14] S. Ferrara and A. Strominger in Strings - 89 (245) (R. Arnowitt, R. Bryan, M.J. Duff, D. Nanopoulos and C. N. Pope, eds World Scientific 1990)
- [15] L. Dixon, V. Kaplunovsky and J. Louis, Nucl. Phys - B329 (1990) 27
- [16] R. Dijkgraaf, E. Verlinde, H. Verlinde, Nucl. Phys. - B352 (1991) 59
- [17] C. Vafa, - " Topological Landau-Ginzburg models " - Prepr. HUPT-90-A064
- [18a] M.K. Gaillard, B. Zumino, Nucl. Phys. - B193 (1981) 221 ; for a review of the method see chapter IV.6 (volume 2) in:
- [18b] L. Castellani, R. D'Auria, P. Fré - " Supergravity and superstrings: a geometric prospective " - World Scientific 1991
- [19a] S. Ferrara, C. Kounnas, D. Lüst and F. Zwirner, - " Duality invariant partition functions and automorphic function superpotentials for (2,2) string compactifications " -, Prepr. CERN-TH 6090/91
- [19b] A. Giveon, E. Rabinovici, G. Veneziano, Nucl. Phys - B322 (1989) 167 ; M. Dine, P. Huet, N. Seiberg, Nucl Phys. - B322 (1989) 301 ; A. Giveon, M. Porrati, Phys. Lett. - 246B (1990) 54
- [20] K. Galicki, Comm. Math. Phys. 108 (1987) 117
- [21] S. Ferrara, C. Kounnas, L. Girardello, M. Porrati Phys. Lett. - B194 (1987) 368
- [22] S. Ferrara, C. Kounnas, L. Girardello, M. Porrati Phys. Lett. - B194 (1987) 358
- [23] S. Ferrara, C. Kounnas, M. Porrati, F. Zwirner Nucl. Phys. - B318 (1989) 75
- [24] A.B. Zamolodchikov and V.A. Fateev, Sov. Phys. JETP 62 (1984) 215; D. Kastor, E. Martinec and S. Shenker, Nucl. Phys. B316 (1989) 590 E.J Martinec, Phys.

- Lett. B217 (1989) 431. C. Vafa and N.P. Warner, Phys. Lett. B218 (1989) 51 J.I. Latorre and C.A. Lütken, Phys. Lett. B222 (1989) 55. S. Cecotti, L. Girardello and A. Pasquinucci, Nucl. Phys. B328 (1989) 701; preprint SISSA 136/89/EP. S. Cecotti, preprint SISSA 69/90/EP.
- [25a] D.Friedan, E. Martinec and S. Shenker, Nucl. Phys. B271 (1986) 93.
- [25b] L. Bonora, M. Matone, F. Toppan and K. Wu, Phys. Lett. B224 (1989) 115; Nucl. Phys. B334 (1990) 717. P.Di Vecchia, F. Pezzella, M. Frau, K. Hornfeck, A. Lerda and S. Sciuto, Nucl. Phys. B333 (1990) 635
- [26] C. Vafa, S. Cecotti Prepr. SISSA-69/91/EP and HUPT-91/A031

TRANSITION FROM CRITICAL TO NON CRITICAL STRING AT HIGH TEMPERATURE

Costas KOUNNAS

*Laboratoire de Physique Théorique de l'Ecole Normale Supérieure
24 rue Lhomond, 75231 Paris, France*

ABSTRACT

We present a method which permits the exact determination of the temperature dependent effective potential. The validity of our method is restricted to the heterotic and type *II* (zero temperature) supersymmetric solutions. At finite temperatures a phase transition occurs towards some new non trivial phases, in which vortices condense.

In the heterotic case we are able to show that the new phase corresponds to a "non critical" superstring in $7 + 1$ dimensions. Also, we show the stability of the new phase, (at least at the one loop level), due to a new type of symmetry, namely, the so-called "space-like supersymmetry".

The exponential growth of the density of massive states in string theory gives rise to a divergence in the partition function of a string gas at a critical temperature ^[1] (Hagedorn temperature T_H). The value of the latter depends only on the (super)-conformal properties of the underlying two-dimensional theory and not on the particular string model ^[2]. Despite earlier speculations on a possible limiting temperature ^[1,2,3], it was soon realized that a phase transition must occur around T_H since a "solitonic" mode becomes tachyonic for temperatures just above T_H ^[4,5].

Many speculations and conjectures are present in the literature about the new vacuum after the phase transition around T_H ^[1-6]. Obviously, in order to analyse this transition and to identify the new phase it is necessary to derive the temperature dependent effective potential of the "T-winding" mode responsible for the phase transition. In ref. [7], it is shown that this is possible in certain cases, namely in all heterotic and type II supersymmetric string solutions. The crucial observation which is explored in this reference is that the D -dimension superstrings at finite temperature look like $(D - 1)$ -dimensional strings with spontaneously broken supersymmetry after a Euclidean rotation has been performed. The use of space time supersymmetry, even in its spontaneously broken version, is strong enough to fix completely the desired effective action of the would-be tachyonic "T-winding" string states. Thus, our result is based on an extension of the recently proposed method of calculating the effective action of some massive string states ^[8] to the case of strings with spontaneously broken supersymmetry ^[7]. Here I will explain briefly this extension and I will present the results of ref. [7].

A convenient way to study strings in finite temperature is to proceed as in field theories; e.g. we continue to the Euclidean space and compactify the "imaginary" time coordinate on a circle of radius R proportional to the inverse temperature ($R = 1/\pi T$). The logarithm of the partition function can then be identified with the corresponding one-loop vacuum graph.

Applying this method to the free string gas, one finds ^[9,5,10]:

$$\log Z[R] = \int_{\mathcal{F}} \frac{d\tau \wedge d\bar{\tau}}{(\Im \tau)^2} \sum_{\substack{\tilde{m}, n \\ a, b}} Z \left[\begin{smallmatrix} n \\ \tilde{m} \end{smallmatrix} \right] (R; \tau) S \left[\begin{smallmatrix} n, a \\ \tilde{m}, b \end{smallmatrix} \right] Z_0 \left[\begin{smallmatrix} a \\ b \end{smallmatrix} \right] (\tau). \quad (1)$$

$Z_0 \left[\begin{smallmatrix} a \\ b \end{smallmatrix} \right]$ is the zero temperature, spin structure dependent partition function, the parameters a and b take on the values 0 or 1 corresponding to antiperiodic or periodic world sheet fermions and a distinguishes space-time bosons ($a = 0$) from fermions ($a = 1$). $Z \left[\begin{smallmatrix} n \\ \tilde{m} \end{smallmatrix} \right] (R; \tau)$ corresponds to the classical contribution of the compactified coordinate (euclidean time), depending on the winding integers $\tilde{m}$ and n :

$$Z \left[\begin{smallmatrix} n \\ \tilde{m} \end{smallmatrix} \right] = \frac{R}{\sqrt{2}} \exp \left[-\frac{\pi R^2}{\sqrt{2}} \frac{|\tilde{m} - n\tau|^2}{\Im \tau} \right]. \quad (2)$$

$S \left[\begin{smallmatrix} n, a \\ \tilde{m}, b \end{smallmatrix} \right]$ is a spin-statistic sign and expresses the different contributions of space-

time bosons and fermions in finite temperature ^[5,10]. In fact, $S \left[\begin{smallmatrix} n, a \\ \tilde{m}, b \end{smallmatrix} \right]$ is fully determined by the spin-statistic requirement as well as by the modular invariance. Indeed, in the field theory limit ($n = 0$), the space-time fermions give a negative contribution for odd $\tilde{m}$. The generalization to $n \neq 0$ is dictated by modular invariance and leads to ^[10]:

$$S \left[\begin{smallmatrix} n, a \\ \tilde{m}, b \end{smallmatrix} \right] = (-)^{a\tilde{m} + bn + \tilde{m}n\delta} . \quad (3)$$

The parameter δ equals one for the heterotic superstring and zero for type *II*; in the latter case, a, b receive contributions both from left and right movers: $a = a_L + a_R$, $b = b_L + b_R$.

The integrand of (1) with S as in eq. (3) is modular invariant and the integration over the complex modulus τ of the world sheet torus is restricted to the fundamental domain $\mathcal{F} \equiv \{|\tau| > 1, \Im\tau > 0, -\frac{1}{2} < \Re\tau \leq \frac{1}{2}\}$.

Performing a Poisson resummation on $\tilde{m}$ one has

$$\sum_{\tilde{m}, n} \mathcal{Z} \left[\begin{smallmatrix} n \\ \tilde{m} \end{smallmatrix} \right] S \left[\begin{smallmatrix} n, a \\ \tilde{m}, b \end{smallmatrix} \right] = (\Im\tau)^{\frac{1}{2}} \sum_{m, n} (-)^{bn} e^{i\pi\tau P_L^2 - i\pi\tau P_R^2} \quad (4)$$

where P_L and P_R are shifted lattice momenta according to the helicity charge a ^[9]:

$$P_{L,R} = \frac{m + \frac{1}{2}a - \frac{1}{2}n\delta}{R} \pm \frac{nR}{2} . \quad (5)$$

The extra sign factor $(-)^{bn}$ reverses the GSO projection in the odd winding number sectors ^[5,10].

The key observation is that string theories with D -dimensional space-time supersymmetry look at finite temperature as if supersymmetry was spontaneously broken in $(D - 1)$ dimensions ^[10,11]. The breaking is induced because of the spin-statistic factor S which gives rise to a spin structure shifting to the lattice momenta as in eq. (5). To make more transparent the correspondence of states between the temperature induced theory and the supersymmetric one (resulting from a one-dimensional toroidal compactification with unshifted momenta, $S \equiv 1$), one needs to enlarge the latter's Lorentzian lattice $\Gamma^{(1,1)}$ to a lattice $\Gamma^{(2,1)}$ ($\Gamma^{(2,2)}$ in type *II* strings) which also includes the helicity operator Q ($Q = Q_L + Q_R$). In fact, up to a redefinition of m in eq. (5), the spin structure a can be identified with the "charge" of the helicity-type operator

$$Q_{L,R} = \oint (\psi^0 \psi^1)_{L,R} ; \quad Q = \frac{a}{2} . \quad (6)$$

One then finds that a state of the $(D - 1)$ supersymmetric theory with quantum numbers (n, m, Q) is in one-to-one correspondence with the temperature induced

$(D-1)$ theory ^[10,11]:

$$\begin{pmatrix} n \\ m \\ Q \end{pmatrix} \longrightarrow \begin{pmatrix} n \\ m + Q \cdot e - \frac{1}{2} n e \cdot e \\ Q - n e \end{pmatrix}. \quad (7)$$

The parameter e equals to $(1,0)$ in the heterotic case and to $(1,1)$ in the type *II* case ($\delta = e \cdot e$). The shifting of Q as in eq. (7) takes account of the change of the GSO projection by the factor $(-)^{bn}$. The transformation (7) corresponds to a particular deformation of the enlarged lattice, defined by the momenta $(Q_L, P_L; Q_R, P_R)$, which preserves the Lorentzian inner products and, thus, satisfies the modular invariance constraints but modifies the masses. In particular, among previously massless states ($m = n = 0$), all bosons remain massless while all fermions acquire a common "supersymmetry-breaking" mass ^[10,11] $m_{3/2} = \frac{Q \cdot e}{R}$. Finally, the requirements that (7) is consistent with the global properties of the world-sheet supercurrent implies a quantization condition on the breaking parameter e and forces it to be an integer.

The fact that the Hagedorn temperature corresponds to the radius in which the states with winding number $n = \pm 1$ become tachyonic for $R < R_H$, can be easily checked from the general mass formula of the "thermal" spectrum ^[4,5,6,7,10]:

$$\text{bosonic string : } m^2(R) \Big|_{\substack{m=0 \\ n=\pm 1}} = -2 + \left(\frac{R}{2}\right)^2 \Rightarrow R_H = 2\sqrt{2}, \quad (8a)$$

$$\text{heterotic string : } m^2(R) \Big|_{\substack{m=\mp 1 \\ n=\pm 1 \\ Q=\pm 1}} = -1 + \left(\frac{R}{2} - \frac{1}{2R}\right)^2 \Rightarrow R_H = \sqrt{2} + 1, \quad (8b)$$

$$\text{type II string : } m^2(R) \Big|_{\substack{m=\mp 2 \\ n=\pm 1 \\ Q_L=Q_R=\pm 1}} = -1 + \left(\frac{R}{2}\right)^2 \Rightarrow R_H = 2. \quad (8c)$$

Due to this observation, it was conjectured that a phase transition occurs at $T \simeq T_H$, of a Kosterlitz-Thouless ^[12] type, in which vortices condense, (*i.e.* winding modes develop non-zero vacuum expectation values), and the new phase is not described by the old conformal theory on the world-sheet. Furthermore, it was argued that the phase transition must be a first order one because the effective potential of the relevant winding mode ω , evaluated perturbatively around $R \sim R_H$, has a trilinear coupling ^[5] $\sigma \omega \omega^*$. Here σ is the dilaton like modulus field whose expectation value corresponds to the change of R and, therefore, of the temperature.

Obviously, these statements cannot be valid in general. First, in all bosonic string solutions there are tachyons even at zero temperature which shows that all of them are unstable. The same is true for the tachyonic non supersymmetric heterotic and type *II* solutions. It is of main interest to examine the existence of

stable vacuum in these cases but it is outside of the scope of the present work. Here, we will restrict ourselves to heterotic and type *II* solutions which at zero temperatures are stable and have space-time supersymmetry. For these theories we will be able to evaluate the exact effective potential of the winding modes which become tachyonic at $T > T_H$. There is a fundamental difference between the heterotic and type *II* case. In the heterotic case, (assuming always that in zero temperature the theory is supersymmetric), only the modes with $(m, n, Q) = \pm(-1, 1, 1)$ become tachyonic^[7] and $m^2(R)$ is negative only in a finite region of R , $\sqrt{2}-1 < R < \sqrt{2}+1$. This is due to the temperature duality^[5,10] $R \rightarrow 1/R$ in the heterotic case. These statements are not valid in the type *II* case. Since our results are universal and do not depend on the particular choice of the vacuum, we consider, as a simple example, a four-dimensional theory (corresponding to a 5-dimensional zero-temperature string model) with a temperature induced spontaneously broken supersymmetry. In this theory, the interactions among the relevant winding modes of the eqs. (8) and the modulus field associated to the temperature-radius exhibit an extended $N \geq 4$ supersymmetry (spontaneously broken). This can be shown at least for all toroidal and orbifold compactifications since then these fields always belong to the untwisted sector. It turns out that it is often possible in a superstring model to determine the effective action of the massless modes^[13] due to the properties of space-time (extended) supersymmetry^[14]. However, in our case, we encounter two difficulties: (i) the winding mode in which we are interested is not massless in the supersymmetric theory ($e = 0$) and (ii) supersymmetry is not exact but spontaneously broken ($e \neq 0$).

To be more specific, consider all 4d heterotic string models obtained by a toroidal compactification of the ten dimensional superstring which have $N = 4$ space-time supersymmetry and a rank 22 gauge group. They correspond to all possible deformations of the $\Gamma^{(6,22)}$ Lorentzian lattice of the compactified momenta. These deformations depend on 6×22 continuous parameters which can be identified with the VeV's of scalar moduli-fields^[15]. For instance, one can continuously deform a $U(1)^6 \times E_8 \times E_8$ model to a $U(1)^6 \times SO(32)$ or any other $N = 4$ solution. To describe this phenomenon in an effective field theory, one must include all fields corresponding to the states that may become massless for special values of the moduli-space. This is possible because all such states can only belong to vector or gravity $N = 4$ supermultiplets and because the full $N = 4$ supergravity theory, up to two derivatives, is uniquely fixed in terms of its gauging^[14], i.e. the structure constants among all gauge bosons (massless and massive matter gauge bosons, and graviphotons). In the recent work of ref. [8] the "stringy" gauging was determined and the corresponding structure constants were shown to be related to the left and right momenta of the Lorentzian lattice compactification. In a Cartan-Weyl basis one has:

$$f_{Ipq} = P_I \delta_{p+q,0} \quad (9a)$$

$$f_{pqr} = \varepsilon(p, q) \delta_{p+q+r,0} \quad (9b)$$

where $I \equiv (i, \alpha)$ labels the Cartan generators ($i = 1, \dots, 6$ for the graviphotons and $\alpha = 1, \dots, 22$ for the matter gauge group) and $P_i \equiv P_{L_i}$ and $P_\alpha \equiv P_{R_\alpha}$ are precisely the left and right components of the momenta respectively. Finally $\varepsilon(p, q)$ is the cocycle factor phase associated with the vertex operators generating the p and q momentum-states. It has been shown ^[8] that the structure constants of eq. (9) verify the Jacobi identity of some infinite dimensional gauge group probably related to a (non-linearly realized) huge string symmetry.

The above method can be generalized when $N = 4$ supersymmetry is spontaneously broken, since the corresponding effective action is still fixed by the gauging which now involves the graviphotons in a non trivial way ^[7]. One must therefore introduce non zero structure constants f_{IJK} proportional to the deformation parameter ϵ and generalize the flat-group gauging in extended supergravity field theories ^[16].

In order to simplify the presentation of our results on the temperature induced supersymmetry breaking, it is convenient to freeze most of the $N = 4$ states at their minima and keep only the relevant ones; these are the multiplets of the two would-be tachyonic winding modes of eq. (8), $\omega_\pm$, the modulus multiplet associated to the temperature-radius, Z , and the dilaton field S , of the $N = 4$ gravitational supermultiplet. The use of the $SU(4)$ classical symmetry in $N = 4$ ^[14] allows us to further restrict the relevant states to four chiral superfields in a $N = 1$ subsector of the initial $N = 4$ (obtained by an appropriate $Z_2 \times Z_2$ projection). The corresponding $N = 1$ supergravity effective action is then described by a Kähler potential K and a superpotential W appearing through the combination ^[17]:

$$G = K + \ln |W|^2. \quad (10)$$

In ref. ^[7] K and W are derived in both heterotic and type II string as functions of the three complex fields: (a) the dilaton and axion field which is always present in 4d theories $[S]$, (b) the complex moduli field $[Z]$ which is associated to the temperature-radius ($\text{real } Z = R$) and (c) the would-be tachyonic field ω with winding numbers $n = \pm 1$. In both, heterotic and type II string K is given by

$$K = -\log(S + \bar{S}) - 2\log(Z + \bar{Z}) - 4\log(1 - \omega\bar{\omega}) \quad (11)$$

while the superpotential W is different

$$\text{heterotic : } W = \sqrt{2} (1 + 2(Z^2 - 2)\omega^2 + \omega^4) \quad (12a)$$

$$\text{type } II : W = \sqrt{2} (1 + 2(Z^2 - 1)\omega^2 + \omega^4). \quad (12b)$$

The knowledge of K and W determines the full scalar potential via the $N = 1$ formulas of ref. ^[17]:

$$V = e^G \left[G_i (G^{-1})^i_j G^j - 3 \right]. \quad (13)$$

This leads to the following expressions for the potential in the heterotic (eq. (12a)) and type II (eq. (12b)) theories:

$$\text{heterotic : } V = \frac{1}{S} \left[\frac{1 - 6Z^2 + Z^4}{Z^2} R^2 + \frac{4 + 4Z^4}{Z^2} R^4 \right] \quad (14a)$$

$$\text{type II: } V = \frac{1}{4S} \left[(Z^2 - 4)R^2 + 4Z^2 R^4 \right]. \quad (14b)$$

For simplicity, in eqs. (14a, b) all fields are chosen real; extension to the imaginary directions complicates the expressions but does not change the results of minimization. We also introduced the field $\Omega = \frac{\omega}{1-\omega^2}$ taking values in the whole real line when ω runs inside the positive kinetic-energy domain $|\omega| < 1$. The kinetic terms of the fields are also determined by the Kähler potential (11):

$$L_{\text{kin}} = \frac{\partial S \partial \bar{S}}{(S + \bar{S})^2} + 2 \frac{\partial Z \partial \bar{Z}}{(Z + \bar{Z})^2} + 4 \frac{\partial \omega \partial \bar{\omega}}{(1 - \omega \bar{\omega})^2}. \quad (15)$$

Note that, as expected, both the kinetic terms and the potential in the heterotic case, are duality invariant under $Z \rightarrow 1/Z$.

In fig. 1, V (heterotic) is plotted as a function of Z and Ω for S fixed; we immediately observe two regions with qualitatively different behavior:

- (i) For $Z > \sqrt{2} + 1$ (or $Z < \sqrt{2} - 1$), there is one minimum at $\Omega = 0$ for arbitrary Z and S , at which the potential vanishes. This corresponds to the low temperature phase where the flat Z -direction is associated with the temperature.
- (ii) For $\sqrt{2} - 1 < Z < \sqrt{2} + 1$ and given S , the quadratic term of the potential (14a) has a negative coefficient and the previous minimum at $\Omega = 0$ becomes a maximum. There is, however, a unique global minimum at the self dual point $Z = 1$ and a non zero VeV for the solitonic mode $\Omega = \pm \frac{1}{2}$. The value of the potential and, therefore, the genus-zero free energy at the new minimum is negative : $V = -\frac{1}{2S}$.

The new minimum turns out to exhibit universal properties: it satisfies the condition $G_i = 0$ for all fields other than the dilaton which implies that "global" supersymmetry remains unbroken in all "matter" multiplets since the goldstino field $\eta \sim e^G G_i \chi^i$ has only a dilatino component (χ^i is the corresponding fermion of the i^{th} supermultiplet). A simple inspection of eq. (13) then shows that the value of the potential is twice the square of the gravitino mass $m_{3/2} = e^G$:

$$V_{\text{min}}(Z = 1, \Omega = \pm 1/2) = -2 m_{3/2}^2; \quad m_{3/2}^2 = \frac{1}{4S} \quad (16)$$

which is a generic property of all supergravity theories with a dilaton field as in (11), if the equations $G_i = 0$, $i \neq S$ admit a non trivial solution. Moreover, note

that, at the self-dual point $Z = 1$, $m_{3/2}^2 = \frac{1}{4S}$ is independent of the value of the winding mode ω (see eqs. (11) and (12a)).

We have not yet considered the minimization with respect to the dilaton field S . The resulting $V_{\min}$ in eq. (16) can be thought of as a tree level cosmological term which is not constant but provides a runaway dilaton potential. Such a term is in general allowed by conformal invariance and defines the non critical strings; its numerical coefficient is indeed proportional to the "central-charge deficit" δc , which gives the deviation from the critical dimension ^[18]. Defining $S = e^{-\phi}$ and rescaling the metric $g_{\mu\nu} = e^{-\phi} G_{\mu\nu}$, one finds the following convenient form for the effective Lagrangian of the metric $G_{\mu\nu}$ and the dilaton ϕ ,

$$L = e^{-\phi} \left[\frac{1}{2} R - \frac{1}{2} (\partial\phi)^2 - V_0 \right] \quad (17)$$

where R is the curvature scalar and $V_0 = -\frac{1}{2}$ from eq. (16). The equations of motion of (17) are identified with the β -functions of the 2d σ -model

$$L^{(2)} = G_{\mu\nu} \partial X^\mu \partial X^\nu + \phi(X) R^{(2)} \quad (18)$$

with $R^{(2)}$ the 2d curvature scalar, provided $V_0 = \frac{1}{4}(d-10)$. We therefore conclude that in the new minimum, the system loses exactly two dimensions ($\delta\hat{c} = -2$) and corresponds to a non critical string in $7+1$ dimensions (the one refers to the Liouville mode (see below)).

Because of the runaway dilaton potential, there are no static solutions in (17). The simplest non static one consists of choosing the σ -model metric flat, $G_{\mu\nu} = \eta_{\mu\nu}$, and the dilaton linear in a space coordinate $\phi = 2QX_0$, where $V_0 = -2Q^2$ ($Q = 1/2$ in our case). The coordinate X_0 can thus be identified with the Liouville mode, in the so-called strong coupling (intermediate) regime, and is described by a 2d conformal field theory based on the (super) energy-momentum tensor (see eq. (18)):

$$T_B = \frac{1}{2} (\partial X_0)^2 + Q \partial^2 X_0 + \frac{1}{2} \psi_0 \partial \psi_0 \quad (19a)$$

$$T_F = \psi \partial X + 2Q \partial \psi \quad (19b)$$

with central charge $\hat{c}_L = 1 + 8Q^2 = 3$. The non trivial dilaton dependence in eq. (18) gives rise to the second term in T_B which is a background charge for the X_0 space coordinate and modifies the spectrum with a uniform positive mass square shift for all bosons by an amount Q^2 ^[18]:

$$\begin{aligned} \delta M_B^2 &= Q^2 = m_{3/2}^2 \\ \delta M_F^2 &= 0 \end{aligned} \quad (20)$$

where the second equality of the first equation follows from eq. (16). Eq. (20) can also be understood from the space-time point of view because of the non trivial dilaton background. In fact, in the representation (17) a common factor $e^{-\phi} = e^{-2QX_0}$ multiplies all Lagrangian terms; one then finds that all fields that are rescaled by a factor e^{QX_0} satisfy the flat-space wave equation with shifted masses as in eq. (20).

A striking property of the new vacuum is the appearance of a spectrum boson-fermion symmetry, as a "space-like" supersymmetry, which is valid at least for all supermultiplets up to spin two for which we can use the $N = 1$ supergravity action. To evaluate the masses before the linear dilaton background is turned on, we use expression (13) with $G_i = 0$, $i \neq S$:

$$\begin{aligned} V_{k\ell} &= e^G \left[G_{ki} (G^{-1})^i_j G_{\ell}^j + (k \leftrightarrow \ell) - 2G_{k\ell} \right] \\ &= 0 \end{aligned} \quad (21a)$$

and

$$V_{k\bar{\ell}} = e^G \left[G_{ki} (G^{-1})^i_j G_{\bar{\ell}}^j - G_{k\bar{\ell}} \right] \quad (21b)$$

where in the summation over i and $\bar{j}$ the dilaton field is excluded. Eqs. (21) relate the boson to the fermion mass matrix by

$$(M_B^2)_{ij} = (M_F^2)_{ij} - m_{3/2}^2 \delta_{ij} \quad (22)$$

with $^{[17]} M_B^2 = G^{-1/2} V G^{-1/2}$ and $M_F^2 = m_{3/2}^2 G^{-1/2} \hat{G} G^{-1/2}$, where G (V) stand for G_{ij} (V_{ij}) matrices and $\hat{G}$ ($\tilde{G}$) for G_{ij} ($G_{\bar{i}\bar{j}}$) ones. Eq. (22) shows that after taking into account the bosonic mass-shift (20) due to the dilaton motion, the mass spectrum of bosonic and fermionic states is identical ! Note that this property is string-independent and also holds in the context of $N = 1$ supergravity with an S -like dilaton field. It suggests that in addition to the known Poincaré and anti-de Sitter supergravity, there is a new supersymmetry grading in the case of the linear dilaton background.

It is very likely that the above symmetry is valid for the entire theory in the new vacuum although our proof was carried only for the lower spin states. In fact, the new phase seems to be closely related to some previously constructed non critical superstrings in the intermediate region $1 < \hat{c} < 9$ ^[19,20]. In ref. ^[20], an internal $N = 2$ world-sheet supersymmetry is used to project out all complex conformal weights which constitute the most serious problem of this region, via a supersymmetric-like GSO projection. Such a $N = 2$ 2d supersymmetry seems to be present in the high- T phase of the heterotic superstring: at first, the initial $N = 2$ internal, formed by the energy momentum tensor T_B^I , the two supercurrents $T_{F_1}^I$, $T_{F_2}^I$ and one $U(1)$ current J^I , is broken by the temperature induced

supersymmetry breaking due to the coupling of the X^t -torus momenta with the helicity-like charges. The resulting shifts are:

$$J^I \rightarrow J^I + 2Q \partial X^t, \quad T_{F_2}^I \rightarrow T_{F_2}^I + 2Q \partial \psi^t \quad \left(Q = \frac{1}{2} \right). \quad (23a)$$

However, in the new minimum the linear dilaton background induces a shifting

$$T_B^I \rightarrow T_B^I + Q \partial^2 X_0, \quad T_{F_1}^I \rightarrow T_{F_1}^I + 2Q \partial \psi_0 \quad (23b)$$

as described in eq. (19), where X_0 is chosen to be the $N = 2$ partner of the Euclidean time coordinate X^t . Eqs. (23b) combined with (23a) restore a $N = 2$ algebra, while the value $Q = 1/2$ corresponds to $\hat{c} = 7$ non critical superstrings.

In concluding, let us point out that the transition we found here is a generalization of the Kosterlitz-Thouless phase transition ^[12] in the heterotic superstring case. The low temperature region corresponds to the spin phase, where the line of fixed points is identified to the Z -flat direction associated to the definition of temperature. As the system is heated up, the vortex configurations become important and at the critical temperature T_H , a phase transition occurs. The high temperature region corresponds to the vortex condensation phase since the winding mode ω gets a non zero VeV. The Z -flat direction disappears because Z is fixed to the self-dual point and temperature's definition becomes meaningless. This is consistent with the fact that the new phase is described by a moving solution with non-trivial dilaton and metric backgrounds. The system is cooled down to zero temperature since the kinetic energy of these backgrounds cancels exactly the negative vacuum energy and the resulting free energy vanishes, at least up to one loop level, as a consequence of the reappearance of "space-like" supersymmetry. The new vacuum is then perturbatively stable. In the new phase, the system loses two dimensions corresponding to a central charge deficit $\delta\hat{c} = -2$, in contrast to the bosonic case where $\delta c = -1$. This is related to the existence of space-time supersymmetry which implies a $N = 2$ on the world-sheet and leads to a complexification of dimensions. In our previous analysis we considered a heterotic string vacuum in four dimensions; in a different dimensionality the coefficient -3 in eqs. (13) and (16) changes, but the value of gravitino mass $m_{3/2}$ changes too, so that V_0 or Q in eq. (17) remain the same. The precise mechanism of the decoupling of the two dimensions at high temperature remains an open and interesting problem demanding further study. Recently, similar results have been obtained in the context of matrix models with a Kosterlitz-Thouless phase transition of the $c = 1$ to the $c = 0$ non critical bosonic string ^[21]. These results suggest that in our case the high- T phase is described by an infinite number of decoupled $7 + 1$ -dimensional non critical superstrings.

The study of V in type II case (eq. (14b)) leads to a different conclusion. There is no duality symmetry in this case ^[5,10], since in the limit $Z \rightarrow 0$ all

fermions become superheavy and one finds a type *II* purely bosonic and tachyonic theory. As a consequence, the minimization with respect to Ω leads to a runaway potential in the Z -direction and no global minimum exists.

Although our interest in this work was concentrated to string thermodynamics, the mechanism we described can be applied if we replace time with an internal coordinate and can be therefore generalized to heterotic and type *II* theories in which supersymmetry is spontaneously broken by use of different operators. Then one may discover non critical strings in a variety of dimensions.

The most interesting physical application of our result is in the cosmology of the early universe. Reversing the order of the process, the transition of non-critical to critical strings could describe a phase transition in the the history of the early universe. In this scenario, the initial phase corresponds to an infinite number of non interacting "eight"-dimensional "cold" universes described by the moving solution of dilaton and metric backgrounds. At the phase transition, all universes are coupled by long range forces, the dilaton motion stops and its kinetic energy is converted into thermal energy. The system is thermalized in a temperature T_H and describes a "ten"-dimensional slowly expanding universe, which is cooling down adiabatically to the $T \sim 0$ supersymmetric phase of the (critical) heterotic string. A series of questions related to the above evolution and its possible cosmological consequences are currently under investigation.

References

- [1] Hagedorn, R., *Nuovo Cimento Supp.* **3**, 147 (1965); Fubini, S. and Veneziano, G., *Nuovo Cimento* **64A**, 811 (1969); Bardakci, K. and Mandelstam, S., *Phys. Rev.* **184**, 1640 (1969); Huang, K. and Weinberg, S., *Phys. Rev. Lett.* **25**, 895 (1970).
- [2] Antoniadis, I., Ellis, J. and Nanopoulos, D.V., *Phys. Lett.* **199B**, 402 (1987); Axenides, M., Ellis, S.D. and Kounnas, C., *Phys. Rev.* **D37**, 2964 (1988).
- [3] Alvarez, E., *Phys. Rev.* **D31**, 418 (1985); Sundborg, B., *Nucl. Phys.* **B254**, 583 (1985); Bowick, M.J. and Wijewardhana, L.C.R., *Phys. Rev. Lett.* **54**, 2485 (1985); Tye, S.H., *Phys. Lett.* **158B**, 388 (1985).
- [4] Sathiapalan, B., *Phys. Rev.* **D35**, 3277 (1987); Kogan, Ya., *JETP Lett.* **45**, 709 (1987); Axenides, M., *Gen. Rel. and Grav.* **21**, 659 (1989).
- [5] Atick, J.J. and Witten, E., *Nucl. Phys.* **B310**, 291 (1988).
- [6] Abrikosov, A.A. and Kogan, Ya., *Sov. Phys. JETP* **69**, 235 (1989); Campbell, B.A., Ellis, J., Kalara, S., Nanopoulos, D.V. and Olive, K.A., *CERN-TH.5833/90* preprint (1990).
- [7] Antoniadis, I. and Kounnas, C., *Phys. Lett.* **261B**, 369 (1991).
- [8] Giveon, A. and Porrati, M., *Phys. Lett.* **246B**, 54 (1990); UCB-PTH-90-39 preprint (1990).
- [9] Polchinski, J., *Commun. Math. Phys.* **104**, 37 (1986); Ditsas, P. and Floratos, E., *Phys. Lett.* **201B**, 49 (1988).
- [10] Kounnas, C. and Rostand, B., *Nucl. Phys.* **B341**, 641 (1990).
- [11] Rohm, R., *Nucl. Phys.* **B237**, 553 (1984); Ferrara, S., Kounnas, C., Porrati, M. and Zwirner, F., *Nucl. Phys.* **B318**, 75 (1989).
- [12] Berezinskii, V.L., *JETP* **34**, 610 (1972); Kosterlitz, M. and Thouless, D., *J. Phys.* **C6**, 1181 (1973); Villain, J., *J. Phys.* **C36**, 581 (1975).
- [13] Ferrara, S., Kounnas, C. and Porrati, M., *Phys. Lett.* **181B**, 263 (1986); Antoniadis, I., Ellis, J., Floratos, E., Nanopoulos, D.V. and Tomaras, T., *Phys. Lett.* **191B**, 96 (1987); Ferrara, S., Girardello, L., Kounnas, C. and Porrati, M., *Phys. Lett.* **192B**, 368 (1987); **194B**, 358 (1987).

- [14] Cremmer, E., Ferrara, S. and Scherk, J., Phys. Lett. **74B**, 61 (1978); De Roo, M., Nucl. Phys. **B255**, 515 (1985); Phys. Lett. **156B**, 331 (1985); Bergshoeff, E., Koh, I.G. and Sezgin, E., Phys. Lett. **155B**, 71 (1985); see also Derendinger, J.-P. and Ferrara, S. in *Spring School on Supergravity and Supersymmetry*, ICTP, Trieste, 1984.
- [15] Narain, K.S., Phys. Lett. **169B**, 41 (1986); Narain, K.S., Sarmadi, M.H. and Witten, E., Nucl. Phys. **B279**, 369 (1987); Antoniadis, I., Bachas, C. and Kounnas, C., Phys. Lett. **200B**, 297 (1988).
- [16] Scherk, J. and Schwarz, J.H., Phys. Lett. **82B**, 60 (1979); Nucl. Phys. **B153**, 61 (1979); Porrati, M. and Zwirner, F., Nucl. Phys. **B326**, 162 (1989).
- [17] Cremmer, E. *et al.*, Nucl. Phys. **B212**, 413 (1979).
- [18] Antoniadis, I., Bachas, C., Ellis, J., and Nanopoulos, D.V., Phys. Lett. **211B**, 393 (1988); Nucl. Phys. **B328**, 117 (1989); Antoniadis, I. and Bachas, C., Ann. de Phys., Colloque n°1, Suppl. n°6, **14**, 11 (1989).
- [19] Gervais, J.-L. and Neveu, A., Phys. Lett. **151B**, 271 (1985); Bilal, A. and Gervais, J.-L., Nucl. Phys. **B284**, 397 (1987); Phys. Lett. **187B**, 39 (1987).
- [20] Kutasov, D. and Seiberg, N., Phys. Lett. **251B**, 67 (1990).
- [21] Gross, D.J. and Klebanov, I.R., Nucl. Phys. **B344**, 475 (1990); PUPT-1210 preprint (1990).

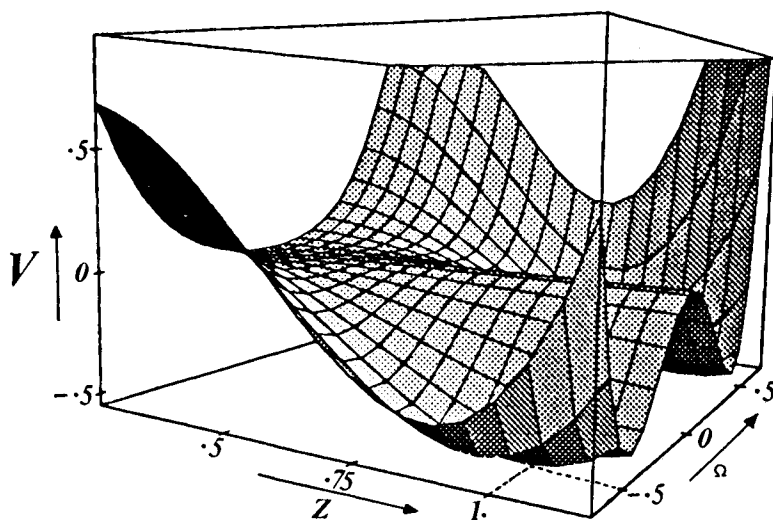


Fig.1 The effective potential $V(Z, \Omega; S = 1)$ of eq. (14a); for $Z < \sqrt{2} - 1$ (or $Z > \sqrt{2} + 1$) the minimum is at $\Omega = 0$, while for $\sqrt{2} - 1 < Z < \sqrt{2} + 1$ there are two degenerate global minima at the self-dual point $Z = 1$ and $\Omega = \pm \frac{1}{2}$.

CANONICAL QUANTIZATION OF $SL(2, \mathbb{R})$ CHERN-SIMONS TOPOLOGICAL THEORIES

CAMILLO IMBIMBO

I.N.F.N., Sezione di Genova

Via Dodecaneso 33, I-16146 Genova, Italy

ABSTRACT

We present a hamiltonian quantization of the $SL(2, \mathbb{R})$ 3-dimensional Chern-Simons theory with fractional coupling constant $k = s/r$ on a space manifold with torus topology in the “constrain-first” framework. We derive the holomorphic multi-valued wave functions on the space of flat $SL(2, \mathbb{R})$ connections on the torus and illustrate their relation to the Kac-Wakimoto characters of the modular invariant representations of $SL(2, \mathbb{R})$ current algebra with level $m = s/r - 2$. The quantum Hilbert space carries an $(s-1)$ -dimensional representation of the modular group with integer and positive fusion rules. The underlying 2-dimensional theory might be regarded as the “square-root” of the Virasoro (r, s) minimal models.

Three-dimensional Chern-Simons topological gauge theory [1] with $SL(2, R)$ as gauge group has attracted considerable interest for various reasons [2]-[5]. There exist arguments relating $SL(2, R)$ Chern-Simons theory to both 2-dimensional [6] and 3-dimensional quantum gravity [7], to the Virasoro discrete series of minimal models with $c < 1$, and, possibly, to some quotient of current algebra models [2],[3]. However, the extension of the Hamiltonian quantization techniques which allowed a non-perturbative solution of Chern-Simons theories with compact gauge groups to theories with non-compact gauge groups is revealed to be problematic [8]. Canonical quantization in the holomorphic “quantize-first” scheme [9]-[11] has been essential for establishing the correspondence between 3-dimensional Chern-Simons theory with compact gauge groups and 2-dimensional current algebras, but this approach is not viable for the real non-compact $SL(2, R)$ due to the lack of a gauge invariant polarization. Analysis based on polarizations which are not gauge invariant [2]-[3] provided some intriguing information about $SL(2, R)$ Chern-Simons theory, but were difficult to carry out at explicit and less formal levels. Recent perturbative computations [5] have stressed the substantially novel features that non-compact gauge groups introduce into the quantization of topological Chern-Simons theories.

In this paper we will present a canonical quantization of Chern-Simons theory with $SL(2, R)$ as gauge group in the “constrain-first” framework [9],[12]. This approach, being gauge invariant *ab initio*, avoids the difficulties of non-gauge invariant polarizations affecting the “quantize-first” method. We will limit ourselves to the case when the “space” manifold Σ is a 2-dimensional torus, which has been sufficient to unravel the underlying 2-dimensional current algebra structure in the compact case. In the “constrain-first” approach, the starting point is the classical gauge invariant phase space $\mathcal{M}$, the space of flat gauge connections on the space manifold Σ . $\mathcal{M}$ is finite-dimensional, so the canonical quantization problem actually has finite number of degrees of freedom. However, the fact that $\mathcal{M}$ is not in general a smooth manifold, makes its quantization rather non-standard. Even in the case of compact gauge groups, $\mathcal{M}$ has singularities of finite order which give rise to important quantum-mechanical effects, such as the “shift” of the central charge in the Sugawara construction for 2-dimensional algebras [12]. When the gauge group is $SL(2, R)$, the singularities of $\mathcal{M}$ are of a more general type, as

we will shortly see: they play a central role in the quantization of the $SL(2, R)$ Chern-Simons theory considered in this work.

Flat $SL(2, R)$ connections on a torus correspond to pairs (g_1, g_2) of commuting $SL(2, R)$ elements, modulo overall conjugation in $SL(2, R)$. g_1 and g_2 represent the holonomies of the flat connections around the two non trivial cycles of the torus. $SL(2, R)$ has a non-trivial Z_2 center and $SO(1, 2) \approx SL(2, R)/Z_2$. Therefore, $\mathcal{M}$ is a four-cover of the space $\mathcal{M}'$ of $SO(1, 2)$ flat connections, since to each $SO(1, 2)$ flat connection correspond four $SL(2, R)$ flat connections whose holonomies differ by elements of the center Z_2 . It is convenient to describe $\mathcal{M}$ in terms of the simpler $\mathcal{M}'$. Let us think of $SL(2, R)$ as the group of unimodular 2×2 real matrices. The basic fact of $\mathcal{M}$ (or $\mathcal{M}'$) is that it is the union of three "sectors"

$$\mathcal{M} = \bigcup_{i=1,2,3} \mathcal{M}_i, \quad (1)$$

where the $\mathcal{M}_i$'s, $i = 1, 2, 3$, are the space of $SL(2, R)$ flat connections whose respective holonomies have two imaginary (and conjugate) eigenvalues ($i = 1$), two eigenvectors with real (and reciprocal) eigenvalues ($i = 2$), and one single eigenvector with unit eigenvalue ($i = 3$).

When $i = 1$, both holonomies can be simultaneously brought by conjugation into the compact $U(1)$ subgroup of $SL(2, R)$. Therefore $\mathcal{M}_1 \approx T^{(1)}$, the two dimensional torus. Let us introduce the real coordinates (θ_1, θ_2) for $\mathcal{M}_1$. In our normalization, the periodic coordinates $\theta_{1,2}$ lie in the unit real interval when the gauge group is $SO(1, 2)$; for $SL(2, R)$, these take values in the enlarged interval of length 2.

For $i = 2$, the holonomies can be conjugated into a diagonal form. However, one can still conjugate diagonal holonomies by an element of the gauge group which permutes the eigenvalues. Therefore, when the gauge group is $SO(1, 2)$, $\mathcal{M}_2 \approx R^{(2)}/Z_2$. If (x, y) are cartesian coordinates on the real plane $R^{(2)}$, the Z_2 action is the reflection around the origin, mapping (x, y) onto $(-x, -y)$. If the gauge group is $SL(2, R)$, $\mathcal{M}_2$ consists of four copies of $R^{(2)}/Z_2$.

Finally, when $i = 3$, holonomies can be conjugated into an upper triangular form with units on the diagonal. Conjugation allows one to rescale the (non-vanishing) elements in the upper right corner by an arbitrary positive number. Thus, $\mathcal{M}_3 \approx S^1$, the real circle. Being odd-dimensional, S^1 cannot be a genuine non-degenerate symplectic space. In fact, the symplectic form on the space of flat connections coming from the Chern-Simons action, when pushed down to $\mathcal{M}_3$ vanishes identically. $\mathcal{M}_3$ represents a “null” direction for the symplectic form of the $SL(2, R)$ Chern-Simons theory, reflecting the indefiniteness of the $SL(2, R)$ Killing form. Since $\mathcal{M}_3$ is a *disconnected* piece of the total phase space $\mathcal{M}$ (or $\mathcal{M}'$), it is consistent to consider the problem of quantizing $\mathcal{M}_1 \cup \mathcal{M}_2$ independently of $\mathcal{M}_3$. After all, modding out by the “null” directions (such as those originated by gauge symmetries), is the common recipe for dealing with degenerate symplectic forms. Hence, in what follows we concentrate on $\mathcal{M}_1 \cup \mathcal{M}_2$, though it is conceivable that the “light-like” sector $\mathcal{M}_3$ merits further investigation.

To summarize, the space of gauge flat connections on the torus (disregarding $\mathcal{M}_3$) looks as follows: a torus ($\mathcal{M}_1$) with planes ($\mathcal{M}_2$) “attached” to it at the points z_s in a discrete set $\mathcal{N} \equiv \mathcal{M}_1 \cap \mathcal{M}_2$, representing flat connections with holonomies in the center of the gauge group. For the $S^0(1, 2)$ case, $\mathcal{N}$ contains a single point, whose $\mathcal{M}_1$ and $\mathcal{M}_2$ coordinates are $(\theta_1^{(s)}, \theta_2^{(s)}) = (0, 0)$ and $(x^{(s)}, y^{(s)}) = (0, 0)$, respectively. When the gauge group is $SL(2, R)$, $\mathcal{N}$ consists of four points, with $(\theta_1^{(s)}, \theta_2^{(s)}) = (\pm 1, \pm 1)$. The $\mathcal{M}_2$ planes are “folded” by the Z_2 reflections around the points in $\mathcal{N}$.

The distinctive feature of classical phase space $\mathcal{M}$ is that it ceases to be a smooth manifold around the points in $\mathcal{N}$. The quantization of the classical phase space $\mathcal{M}$ involves considering smooth functions or smooth sections of appropriate line bundles on $\mathcal{M}$, raising the question of the meaning of “smooth” sections on a non-smooth manifold such as $\mathcal{M}$. Our strategy is to consider first the smooth, non-compact manifold $\mathcal{M}/\mathcal{N}$ obtained by deleting the singular points in $\mathcal{N}$. $\mathcal{M}/\mathcal{N}$ consists of two disconnected smooth components, $\mathcal{M}_1/\mathcal{N}$ and $\mathcal{M}_2/\mathcal{N}$. We will then consider quantizations of $\mathcal{M}_1/\mathcal{N}$ and $\mathcal{M}_2/\mathcal{N}$ which admit sections that can be “glued” at the points in $\mathcal{N}$. The final Hilbert space will be the span of those “glued” sections. Our “intuitive” approach could conceivably be substantiated with more rigorous methods of algebraic geometry.

We will perform the quantization of $\mathcal{M}_1/\mathcal{N}$ and $\mathcal{M}_2/\mathcal{N}$ in the holomorphic scheme [9]-[11] since, as is familiar from the study of the compact Chern-Simons theory [1], $\mathcal{M}$ admits a natural family $\mathcal{T}$ of Kähler polarizations [13]. $\mathcal{T}$ is the Siegel upper complex plane, because the choice of a complex structure on the 2-dimensional space manifold Σ induces a complex structure on the space of connections on Σ and, by projection, on $\mathcal{M}$. For $\tau \in \mathcal{T}$, let us introduce holomorphic coordinates on both $\mathcal{M}_1/\mathcal{N}$

$$z = \theta_1 + \tau\theta_2, \quad \bar{z} = \theta_1 + \bar{\tau}\theta_2, \quad (\theta_1, \theta_2) \in \mathcal{M}_1/\mathcal{N} \quad (2)$$

and $\mathcal{M}_2/\mathcal{N}$

$$z = x + \tau y, \quad \bar{z} = x + \bar{\tau} y, \quad (x, y) \in \mathcal{M}_2/\mathcal{N}. \quad (3)$$

Then the symplectic form which descends from the Chern-Simons action with charge k

$$S = \frac{k}{4\pi} \int_{\Sigma \times R^1} \langle A, dA + \frac{1}{3} A \wedge A \rangle \quad (4)$$

can be written both on $\mathcal{M}_1$ and $\mathcal{M}_2$ in the coordinates systems (2) and (3) as follows:

$$\omega = \frac{ik\pi}{2\tau_2} dz \wedge d\bar{z} \equiv i\bar{\partial}\partial K, \quad \tau_2 \equiv \text{Im}\tau \quad (5)$$

where K is the Kähler potential which we choose to be

$$K = \frac{k\pi}{4\tau_2} (z - \bar{z})^2. \quad (6)$$

In the context of Kähler quantization, the Hilbert space of quantum states is the span of square integrable holomorphic sections of a holomorphic line bundle with hermitian structure whose curvature two-form is the symplectic form ω in (5).

The quantization of $\mathcal{M}_2/\mathcal{N}$ is rather straightforward. Since $\mathcal{M}_2/\mathcal{N}$ is not simply connected, the holomorphic wave functions $\psi(z)$ can acquire an arbitrary phase $e^{2\pi i\vartheta_i}$ when moving around the singular points $z_i = 0$ of $\mathcal{N}$. The Böhm-Aharonov phases $e^{2\pi i\vartheta_i}$ should be regarded as free parameters of the quantization.

A further two-fold ambiguity of the $\mathcal{M}_2/\mathcal{N}$ quantization stems from the fact that the gauge invariant $\mathcal{M}_2/\mathcal{N}$ is the quotient of the complex plane (with the origin deleted) by the action of the reflection around the origin. Thus, physical wave functions should be invariant under the action of the unitary operator implementing the reflection around the origin. Since there are two ways of implementing reflections according to the “intrinsic” parity of the wave functions, one concludes that the wave functions on each of the four “sheets” of $\mathcal{M}_2/\mathcal{N}$ are

$$\psi^{(\vartheta, \pm)}(z) = z^\vartheta \chi^{(\pm)}(z), \quad (7)$$

where $\chi^{(\pm)}(z)$ is holomorphic, even (odd) around the origin, and each choice of $(\theta, \pm)$ is associated to a different quantum Hilbert space $\mathcal{H}_{\mathcal{M}_2/\mathcal{N}}^{(\theta, \pm)}$.

Let us now turn to $\mathcal{M}_1/\mathcal{N}$. We need to look for a holomorphic line bundle $\mathcal{L}_o^{(k)}$ on $\mathcal{M}_1/\mathcal{N}$ with hermitian structure having the curvature two-form given by Eq. (5). Since $\mathcal{M}_1/\mathcal{N}$ is non-compact, there is no quantization condition on k associated to the integer-valuedness of the Chern class of the symplectic form, unlike in the case of Chern-Simons theories with compact gauge groups. When k is irrational one expects an infinite-dimensional Hilbert space of holomorphic sections: an interesting possibility, which we do not pursue here. We focus rather on the connections between $SL(2, R)$ Chern-Simons theory and 2-dimensional rational conformal field theories. Therefore, we restrict ourselves henceforth to the case of k *rational*:

$$k = s/r, \quad (8)$$

with s and r integers, relatively prime, and r chosen to be positive.*

The crucial difference between quantum mechanics on the non-compact $\mathcal{M}_1/\mathcal{N}$ and on the compact torus $\mathcal{M}_1$ originates from the fact that the homotopy group $\pi_1(\mathcal{M}_1/\mathcal{N})$ is *non-abelian*:

$$ab = ba\delta, \quad [a, \delta] = [b, \delta] = 0, \quad (9)$$

* Abelian Chern-Simons theories with fractional coupling constant have been studied, independently and in a different context, in [14].

where a and b are the non-trivial cycles of the compact torus and $\delta = \prod_i \delta_i$ is the product of the cycles δ_i surrounding the singularities z_i in $\mathcal{N}$. In this case, quantum states are represented by *multi-valued* wave-functions $\Psi(z) = (\psi^\alpha(z))$, $\alpha = 0, 1, \dots, q-1$, transforming in some irreducible unitary, q -dimensional representation of the homotopy group $\pi_1(\mathcal{M}_1/\mathcal{N})$. Let us consider a basis for such representation which diagonalizes the δ_i 's. For the representation to be finite-dimensional, the phases representing the δ_i 's must be *rational*. Moreover, we take all δ_i 's to be the same, since we require that modular transformations (which mix the singular points in $\mathcal{N}$) act on the Hilbert space of wave functions. Thus, $\delta = \exp(-2\pi i p/q)$ with p integer, coprime with q . On the other hand, $\Psi(z)$ should be a section of $\mathcal{L}_o^{(k)}$ on $\mathcal{M}_1/\mathcal{N}$, implying that in the trivialization corresponding to (6), $\Psi(z)$ has the periodicity properties of theta functions with fractional "level" k :

$$\Psi(z + 2m + 2n\tau) = \exp(-2\pi i k \tau n^2 - 2\pi i k z n) a^m b^n \Psi(z), \quad (10)$$

where a and b are $q \times q$ unitary matrices which provide a representation of homotopy relations (9). Note that on the compact torus $\mathcal{M}_1$, a and b would be one-dimensional phases and the consistency (cocycle) condition for the transition functions in (10) would require k to be an integer [9]-[11]. In our case, the consistency condition coming from (9) imposes

$$2k = p/q. \quad (11)$$

In what follows, we choose r to be odd for the sake of concreteness; thus (11) means that $r = q$ and $p = 2s$. An explicit matrix representation for a and b is :

$$\begin{aligned} (a)_{\alpha\beta} &= e^{-2\pi i (2k)\alpha} \delta_{\alpha,\beta} \\ (b)_{\alpha\beta} &= \delta_{\alpha,\beta+1}, \quad \alpha, \beta = 0, 1, \dots, r-1. \end{aligned} \quad (12)$$

All other r -dimensional representations of (9) are obtained from (12) by means of rescaling

$$\begin{aligned} a &\rightarrow \exp(2\pi i \vartheta_a) a \\ b &\rightarrow \exp(2\pi i \vartheta_b) b. \end{aligned} \quad (13)$$

Phases $e^{2\pi i \vartheta_{a,b}}$ which differ by r -roots of the unity give rise to equivalent representations of (9). Modular invariance requires $e^{(2\pi i r \vartheta_{a,b})} = 1$, similarly to the compact case [9]-[11]. One can therefore assume (12).

In geometric quantization, in order to implement canonical transformations which do not leave the polarization invariant (such as modular transformations), the wave functions $\Psi_\tau(z)$ are also regarded as dependent on the polarization $\tau \in \mathcal{T}$. The τ dependence is determined by the requirement that quantum Hilbert spaces $\mathcal{H}_\tau$ relative to different τ 's be unitarily equivalent with respect to the hermitian forms

$$\left(\Psi_\tau^{(1)}, \Psi_\tau^{(2)}\right) = \int_{\mathcal{M}_1} dz d\bar{z} \tau_2^{-1/2} e^{-\frac{k\pi}{4\tau_2}(z-\bar{z})^2} \Psi_\tau^{(1)}(\bar{z})^* \Psi_\tau^{(2)}(z) \quad (14)$$

associated to the Kähler structure (5). This implies that the wave functions $\Psi_\tau(z)$ should be *parallel* with respect to a flat, unitary connection on the vector bundle with base $\mathcal{T}$ and fibers $\mathcal{H}_\tau$ [11]. Therefore, an orthonormal $2s$ -dimensional basis for the r -valued parallel sections of the quantization of $\mathcal{M}_1/\mathcal{N}$ is given by [11],[12]:

$$(\Psi_N(\tau; z))^\alpha \equiv \psi_N^\alpha(\tau; z) = \theta_{rN+2\alpha s, rs}(\tau; z/r), \quad N = 0, 1, \dots, 2s-1, \quad (15)$$

where the $\theta_{m,rs}(\tau; z)$ (m integer modulo $2rs$) are level rs $SU(2)$ theta functions [15].

Among classical canonical transformations, reflections w around the singular points in $\mathcal{N}$ are of special interest for our purposes:

$$w : z \rightarrow -z. \quad (16)$$

In the multi-component description of the holomorphic wave-functions on $\mathcal{M}_1/\mathcal{N}$, w will be implemented on the Hilbert space as follows:

$$U(w) : \Psi(z) \rightarrow W\Psi(-z), \quad (17)$$

where W is an $r \times r$ unitary matrix acting on the "internal" indices α . W is determined by the condition that (17) defines an automorphism of the line bundle $\mathcal{L}_o^{(k)}$

$$W a^m b^n = a^{-m} b^{-n} W. \quad (18)$$

In the representation (12) the solution of (18) is:

$$(W)_{\alpha,\beta} = \delta_{\alpha,-\beta}. \quad (19)$$

The Hilbert space $\mathcal{H}_{\mathcal{M}_1/\mathcal{N}}$ spanned by the sections (15) splits under the action of $U(w)$ into “even” and “odd” subspaces $\mathcal{H}_{\mathcal{M}_1/\mathcal{N}}^{\pm}$, with orthonormal parallel basis given by the following combinations of theta functions:

$$(\Psi_N^{\pm})^{\alpha} \equiv \psi_N^{\alpha,(\pm)}(\tau; z) = \theta_{rN+2\alpha s,rs}(\tau; z/r) \pm \theta_{-rN+2\alpha s,rs}(\tau; z/r). \quad (20)$$

Were we simply trying to quantize $\mathcal{M}_1/\mathcal{N}$, we would keep both the even and the odd sector since canonical transformations (16) in the $\mathcal{M}_1$ sector do *not* correspond to gauge transformations of the original Chern-Simons $SL(2, R)$ theory. However, we really want to quantize the union $\mathcal{M}_1 \cup \mathcal{M}_2$. In the $\mathcal{M}_2$ sector, reflections (16) do correspond to gauge symmetries of the Chern-Simons theory and we must choose between the “even” and the “odd” wave functions, as stated in (7). Since wave functions (7) in $\mathcal{H}_{\mathcal{M}_2/\mathcal{N}}$ have to be “glued” at the points in $\mathcal{N}$ to those in $\mathcal{H}_{\mathcal{M}_1/\mathcal{N}}$, their respective monodromies around the singular points should match, requiring that $\exp(2\pi i \vartheta) = \exp(2\pi i s/2r)$. Moreover, wave functions in $\mathcal{H}_{\mathcal{M}_1/\mathcal{N}}$ and in $\mathcal{H}_{\mathcal{M}_2/\mathcal{N}}$ should have the same behaviour under reflections around the singular points. In a single-component description, “odd” (“even”) wave functions on $\mathcal{M}_1/\mathcal{N}$ near the singular point z_i behave as $\sim (z - z_i)^{(s/2r)}(z - z_i)$ ($\sim (z - z_i)^{(s/2r)}$), and therefore can be glued to wave functions (7) on $\mathcal{M}_2/\mathcal{N}$, with $\chi^{(-)}(z) = z$ ($\chi^{(+)}(z) = 1$). In conclusion, the phase space $\mathcal{M}_1 \cup \mathcal{M}_2$ admits two inequivalent quantizations, with Hilbert spaces isomorphic to $\mathcal{H}_{\mathcal{M}_1/\mathcal{N}}^{\pm}$. A similar ambiguity was present in the $SU(2)$ case [12], but it was the “odd” quantization which was related to 2-dimensional conformal field theories for generic k . In fact, only the “odd” projection gives positive integer fusion rules for k generic, suggesting that this is the quantization of the Chern-Simons theory on the torus which generalizes, in some appropriate sense, to higher genus space manifolds [11]. In our case as

well, “odd” quantization gives positive, integer fusion rules for generic k , as we will shortly see, though we do not yet know its 2-dimensional interpretation. From now on, we therefore concentrate on $\mathcal{H}_{\mathcal{M}_1/\mathcal{N}}^- \equiv \mathcal{H}_{CS}$.

Wave-functions in $\mathcal{H}_{CS}$ are related to the Kac-Wakimoto characters [16] of irreducible, modular invariant representations of $SL(2, R)$ current algebra with fractional central charge m

$$m \equiv t/u = s/r - 2, \quad (21)$$

i.e., with $u = r$ and $s = 2u + t$. The Kac-Wakimoto characters are defined as follows:

$$\chi_{j(N', \alpha'); m}(z, \tau) = \text{tr}_{\mathcal{H}_{j, m}} e^{2\pi i \tau L_0 + 2\pi i z J_0^3}, \quad (22)$$

where $\mathcal{H}_{j, m}$ is the highest weight irreducible representation of $SL(2, R)$ current algebra with level m and spin j . $j = j(N', \alpha')$ ranges over the following set:

$$j = 1/2(N' - \alpha'(m + 2)), \quad N' = 1, 2, \dots, 2u + t - 1, \quad \alpha' = 0, 1, \dots, u - 1. \quad (23)$$

One can check that the orthonormal wave functions (20) of $\mathcal{H}_{CS}$ can be written in terms of the Kac-Wakimoto characters:

$$\frac{\psi_N^{\alpha, (-)}(\tau; z)}{\Pi(\tau; z)} = \begin{cases} \chi_{j(N, 2\alpha)} & \text{if } \alpha \in \{0, 1, \dots, (r-1)/2\} \\ \chi_{j(s-N, 2\alpha-r)} & \text{if } \alpha \in \{(r+1)/2, \dots, r-1\}, \end{cases} \quad (24)$$

where $\Pi(\tau; z)$ is the Kac-Wakimoto denominator:

$$\Pi(\tau; z) = \theta_{1,2}(\tau, z) - \theta_{-1,2}(\tau, z). \quad (25)$$

$\Pi(\tau; z)$ is holomorphic and non-vanishing on $\mathcal{M}_1/\mathcal{N}$. Therefore, the wave functions $\Psi_N^{(-)}(\tau; z)$ and the wave functions

$$\Psi'_N(\tau; z) = \frac{\Psi_N^{(-)}(\tau; z)}{\Pi(\tau; z)}$$

appearing in (24), describe equivalent sections of $\mathcal{L}_o^{(k)}$ on $\mathcal{M}_1/\mathcal{N}$. These are related by the change of trivialization of the line bundle on $\mathcal{M}_1/\mathcal{N}$ associated to the Kähler transformation

$$K \rightarrow K + f(z) + f^*(\bar{z}), \quad (26)$$

with $f(z) = -\ln(\Pi(\tau; z))$. (26) introduces in the scalar product (14) a factor which has been interpreted as the jacobian of a certain change of integration variables in the path integral formulation of the Chern-Simons theory [9]. (24) implies that to each quantum state emerging out of the quantization of the $SL(2, R)$ Chern-Simons theory, there corresponds a *multiplet* of $SL(2, R)$ current algebra characters, rather than one single character as in the $SU(2)$ case. One does not expect, therefore, that the hypothetical 2-dimensional theory underlying 3-dimensional $SL(2, R)$ Chern-Simons theory has full $SL(2, R)$ current algebra symmetry. One might speculate that such a theory could be obtained from some coset construction of $SL(2, R)$ current algebra, though not a standard coset as will become apparent from the analysis of the modular transformation properties which follows.

Let us use s and t to denote the canonical transformations of the classical phase space $\mathcal{M}_1/\mathcal{N}$ which generate the modular group $SL(2, Z)$ of the torus

$$\begin{aligned} s : (\tau, z) &\rightarrow (-1/\tau, z/\tau) \\ t : (\tau, z) &\rightarrow (\tau + 1, z), \end{aligned} \quad (27)$$

and satisfy the relations:

$$s^2 = 1, \quad (st)^3 = 1. \quad (28)$$

s and t will be represented on the space of the multi-valued wave functions $(\Psi(z))^\alpha \equiv \psi^\alpha(\tau; z)$ on $\mathcal{M}_1/\mathcal{N}$ by means of unitary operators $U(s)$ and $U(t)$:

$$\begin{aligned} U(s) : \Psi(\tau; z) &\rightarrow (S^{-1}\Psi)(-1/\tau; z/\tau) \\ U(t) : \Psi(\tau; z) &\rightarrow (T^{-1}\Psi)(\tau + 1; z), \end{aligned} \quad (29)$$

where $S \equiv (S)_{\alpha\beta}$ and $T \equiv (T)_{\alpha\beta}$ are unitary $r \times r$ matrices acting on the “internal” indices and providing a representation of the modular group. In order to define

an automorphism of the holomorphic line bundle $\mathcal{L}_o^{(k)}$, S and T should commute with the $\mathcal{L}_o^{(k)}$ transition functions:

$$\begin{aligned} a^m b^n T &= T a^{m+n} b^n \\ a^m b^n S &= S a^n b^{-m}. \end{aligned} \quad (30)$$

Eqs. (30) and (12) give for T and S :

$$\begin{aligned} (T)_{\alpha\beta} &= \delta_{\alpha,\beta} e^{2\pi i \frac{r}{s} \alpha^2 - 2\pi i \theta(2s;r)/3} \\ (S)_{\alpha\beta} &= 1/\sqrt{r} e^{2\pi i \frac{2s}{r} \alpha\beta}, \end{aligned} \quad (31)$$

where the phase $\theta(2s;r)$ is determined from the $SL(2, Z)$ relation, $(ST)^3 = 1$. Thus:

$$e^{2\pi i \theta(r;s)} = 1/\sqrt{s} \sum_{n=0}^{s-1} e^{2\pi i \frac{r}{2s} n^2} \quad (32)$$

(r and s coprime and rs even). It is well-known that $\theta(1;2s) = 1/8 \bmod 1$, agreeing with the fact that the conformal central charge of a free compactified 2-dimensional scalar field is one. From definition (32) it follows that:

$$\theta(r';s) \equiv \theta(r;s) \bmod 1, \text{ if } r' \equiv rn^2 \bmod s \quad (33)$$

for n integer. An explicit formula for $\theta(r;s)$ can be derived for generic r by considering the modular properties of the following theta function:

$$\theta(\tau) \equiv \sum_{n \in Z} e^{-i\pi n^2 \tau}. \quad (34)$$

In fact, one can show that the phase $e^{2\pi i \theta(r;s)}$ is related to the multiplier system of the modular form (34):

$$\theta(\alpha(\tau)) = e^{-2\pi i \theta(r;s)} \left(\frac{s\tau + r}{i} \right)^{1/2} \theta(\tau), \quad (35)$$

where $\alpha(\tau) \equiv \frac{a\tau+b}{s\tau+r}$ is a modular transformation belonging in Γ_θ , the subgroup of $SL(2, Z)$, with as and br even. From (35) one derives

$$\theta(r; s) = 1/2 S(r; s) - S(s + r; 2s), \quad (36)$$

where $S(r; s)$ is the *Dedekind symbol* [17]:

$$S(r; s) = \sum_{n=1}^{s-1} ((n/s - 1/2)) ((nr/s - 1/2)), \quad (37)$$

with $((x)) \equiv x$ modulo integers, $-1/2 \leq ((x)) < 1/2$. It also follows from (35) that

$$e^{8\pi i \theta(r; s)} = -1, \quad (38)$$

i.e., that the allowed values for $\theta(r; s)$ are $\pm 1/8$ and $\pm 3/8 \pmod{1}$.

The representations of the modular group acting on the quantum Hilbert spaces $\mathcal{H}_{\mathcal{M}_1/N}$ and $\mathcal{H}_{CS}$ can now be derived from the modular properties of level rs theta functions (15), (20) and from representation (31) acting on the “internal” indices. In order to make our formula more transparent, let us introduce the following $2s$ -dimensional matrix representation $\mathbf{R}^{(r; 2s)}$ of the modular group:

$$\begin{aligned} T_{N,M} &= e^{2\pi i \frac{r}{4s} N^2 - 2\pi i \theta(r; 2s)/3} \delta_{N,M} \\ S_{N,M} &= 1/\sqrt{2s} e^{2\pi i \frac{r}{2s} NM} \quad N, M = 0, 1, \dots, 2s-1. \end{aligned} \quad (39)$$

For $r = 1$, this is the representation of the modular group associated to the characters of a 2-dimensional scalar field compactified on a circle of radius $R^2 = m/n$ with $2s = mn$. Note that $S^2 = C$, with $(C)_{N,M} = \delta_{N,-M}$ being the “charge conjugation” matrix. Since C commutes with the matrices in (39), $\mathbf{R}^{(r; 2s)}$ decomposes into two representations $\mathbf{R}_{\pm}^{(r; 2s)}$, even and odd under C :

$$\mathbf{R}^{(r; 2s)} = \mathbf{R}_{+}^{(r; 2s)} \oplus \mathbf{R}_{-}^{(r; 2s)}, \quad (40)$$

where $\mathbf{R}_{\pm}^{(r; 2s)}$ is $s \pm 1$ dimensional. Since $S_-^2 = -1$, it is convenient to define a $\tilde{\mathbf{R}}_-^{(r; 2s)}$ by $\tilde{S}_- = -iS_-$ and $\tilde{T}_- = iT_-$, such that the charge conjugation matrix is equal to the identity. ($\tilde{\mathbf{R}}^{(1; 2s)}$ is the representation associated to the characters of the integrable representations of the $SU(2)$ current algebra of level $s - 2$.) Eqs.

(15),(20),(31) and (40) imply that the modular representations acting on $\mathcal{H}_{\mathcal{M}_1/N}$ and $\mathcal{H}_{CS}$ are :

$$\begin{aligned}\mathbf{R}_{\mathcal{H}_{\mathcal{M}_1/N}}^{(s/r)} &= \mathbf{R}^{(r;2s)} \\ \mathbf{R}_{CS}^{(s/r)} &= \tilde{\mathbf{R}}_-^{(r;2s)}.\end{aligned}\quad (41)$$

The spectrum of conformal dimensions and the central charge of the hypothetical 2-dimensional conformal field theory underlying the 3-dimensional $SL(2, R)$ Chern-Simons theory which follows from (41) is:

$$\begin{aligned}c &= 2 - 6r/s + 8\theta(r; 2s) \mod 8 \\ h_p &= \frac{r}{4s}(p^2 - \bar{p}^2) \mod 1 \quad p = 1, 2, \dots, s-1,\end{aligned}\quad (42)$$

where $\bar{p} \in \{1, 2, \dots, s-1\}$ corresponds to the conformal block of the identity operator. Note that for $r = 1$ (and $\bar{p} = 1$), the spectrum and central charge in (42) are those of $SU(2)$ current algebra with level $l = s - 2$ and spin $j = (p - 1)/2$.

Eqs. (24) and (31) relate the modular representation $\mathbf{R}_{CS}^{(s/r)}$ to the representation $\mathbf{R}_{KW}^{(s/r)}$ associated with the Kac-Wakimoto current algebra representations (22):

$$\mathbf{R}_{KW}^{(s/r)} = \mathbf{R}^{(2s;r)} \otimes \mathbf{R}_{CS}^{(s/r)}.\quad (43)$$

In [18] a connection was discovered between the Kac-Wakimoto characters and the Rocha-Caridi characters of the $c < 1$ Virasoro minimal models. Together with our result (43), this allows us to establish the following relation between $\mathbf{R}_{CS}^{(s/r)}$ and the representation $\mathbf{R}_{Vir}^{(r;s)}$ associated to the (r, s) minimal model of the Virasoro discrete series:

$$\mathbf{R}_{Vir}^{(r;s)} = \mathbf{R}_{CS}^{(r/4s)} \otimes \mathbf{R}_{CS}^{(s/r)}.\quad (44)$$

Eq. (44) tells us, loosely speaking, that the hypothetical 2-dimensional conformal field theory corresponding to $SL(2, R)$ Chern-Simons topological theory with fractional charge $k = s/r$ can be regarded as the “square root” of the (r, s) minimal model of the Virasoro discrete series.

The algebraic data needed to reconstruct a 2-dimensional rational conformal field theory is not exhausted by the modular representation of the conformal blocks of the genus one partition function. The representations of the modular group associated with conformal field theories, together with the braid matrices, should satisfy a set of rather restrictive identities, known as “the polynomial equations” [19]. In the Chern-Simons framework, the derivation of the braid matrices would require the solution of the theory on a space-manifold with the topology of a sphere with punctures, a problem which we did not address here. The polynomial equations imply certain necessary, but not sufficient, conditions for the representation of the modular group of the conformal blocks of the identity operator on the torus. The most celebrated among these conditions, due to E. Verlinde [20], requires that the numbers N_{ijk} , defined in terms of the modular matrix S as

$$N_{ijk} = \sum_n \frac{S_{in} S_{jn} S_{kn}}{S_{0n}}, \quad (45)$$

be positive and integer, since they are interpreted as the *fusion rules* of the conformal field theory. From the expression for $S_{CS}^{(s/r)}$, derivable from (39)-(41), one obtains the same (positive and integer) fusion rules as for the $SU(2)$ Wess-Zumino model of level $s - 2$:

$$N_{ijk}^{CS} = N_{ijk}^{SU(2)_{s-2}}. \quad (46)$$

(46) can be regarded as the “square-root” of the Virasoro minimal models fusion rules, in agreement with (44).

Modular invariance of 4-point correlation functions of primary operators on the sphere gives rise to another necessary condition for the h_p 's in (42) to be the spectrum of dimensions of a conformal field theory [21]. This condition requires that

$$(\alpha_i \alpha_j \alpha_k \alpha_l)^{N_{ijkl}} = \prod_r \alpha_r^{N_{ijkl,r}}, \quad (47)$$

where $\alpha_i = \exp(2\pi i h_i)$ and

$$N_{ijkl,r} = N_{ijr} N_{klr} + N_{jkr} N_{ril} + N_{ikr} N_{rjl}.$$

One can verify that (47) is indeed satisfied by the dimensions in (42) and the fusion rules in (46).

Unlike in the compact case, the geometric quantization of $SL(2, R)$ Chern-Simons theory does not provide the explicit functions of Teichmüller space which transform according to the representation $\mathbf{R}_{CS}^{(s/r)}$ of the modular group and which could be identified with the Virasoro characters of an underlying 2-dimensional conformal theory. Eqs. (24) and (43) say that the 2-dimensional “object” underlying the $SL(2, R)$ Chern-Simons topological theory is related to 2-dimensional $SL(2, R)$ current algebras. However, since $\mathbf{R}^{(2s;r)}$ appears in (43) (rather than the representation $\mathbf{R}^{(1;r)}$ associated to abelian current algebras) one cannot conclude that the 2-dimensional interpretation of $SL(2, R)$ Chern-Simons theory is a standard coset of $SL(2, R)$ current algebra. Actually, although conditions (45) and (47) are both satisfied, this 2-dimensional “object”, at least for generic k , might belong to a class of quantum field theories more general than the conformal family — a class of theories for which concepts like holomorphic blocks, modular invariance and fusion rules would be still be meaningful. It is tempting to speculate that the 2-dimensional theories associated to $SL(2, R)$ Chern-Simons theories are related to quantum deformations of $SL(2, R)$ Kac-Moody algebras with quantum parameter $q = e^{2\pi i/r}$, to which several concepts of conformal field theories extend.

ACKNOWLEDGEMENTS

I would like to thank A. Schwimmer for discussion and criticism offered throughout the course of this work and R. Jengo for illustrating to me the point of view of [14]. Stimulating conversations with V.G. Kac, M. Wakimoto and S. Mukhi are also acknowledged. I am grateful to the organizers of the “Spring Workshop on String Theory”, Trieste, April 1991, for their invitation to speak.

REFERENCES

- [1] E. Witten, *Comm. Math. Phys.* **121** (1988) 351.
- [2] H. Verlinde, Princeton Preprint, PUTP-89/1140.
- [3] E. Verlinde and H. Verlinde, Princeton Preprint, PUTP-89/1149.
- [4] E. Witten, IAS Preprint, IASSNS-HEP 89/38.
- [5] Dror Bar-Natan and E. Witten, IAS Preprint, IASSNS-HEP-91/4.
- [6] V.G. Knizhnik, A.M. Polyakov and A.B. Zamolodchikov, *Mod. Phys. Lett.* **A3** (1988) 819.
- [7] E. Witten, *Nucl. Phys.* **B311** (1988/9) 46; *Nucl. Phys.* **B323** (1989) 113.
- [8] E. Witten, IAS Preprint IASSNS-HEP-89/65, to appear in *Comm. Math. Phys.*
- [9] S. Elitzur, G. Moore, A. Schwimmer and N. Seiberg, *Nucl. Phys.* **B326** (1989) 108.
- [10] M. Bos and V.P. Nair, *Phys. Lett.* **B223** (1989) 61; *Int. J. Mod. Phys.* **A5** (1990) 959.
- [11] S. Axelrod, S. Della Pietra and E. Witten, IAS Preprint IASSNS-HEP-89/57.
- [12] C. Imbimbo, *Phys. Lett.* **B258** (1991) 353.
- [13] N. Woodhouse, "Geometric Quantization", Oxford University Press, Oxford 1980.
- [14] R. Jengo and K. Lechner, SISSA Preprint, SISSA-182-90-EP.
- [15] V. G. Kac, "Infinite Dimensional Lie Algebras", Cambridge 1985.
- [16] V.G. Kac and M. Wakimoto, *Proc. Nat. Acad. Sci.* **85** (1988) 4956.
- [17] See, for example: S. Lang, "Introduction to Modular Forms", Springer-Verlag, New York 1976.
- [18] S. Mukhi and S. Panda, *Nucl. Phys.* **B338** (1990) 263.
- [19] G. Moore and N. Seiberg, *Phys. Lett.* **B212** (1988) 451; *Comm. Math. Phys.* **123** (1989) 177.
- [20] E. Verlinde, *Nucl. Phys.* **B300** (1988) 360.
- [21] C. Vafa, *Phys. Lett.* **B206** (1988) 421.

SINGULAR VECTORS IN VIRASORO VERMA MODULES

by

Michel Bauer

*Service de Physique Théorique[†] de Saclay
CE-Saclay, F-91191 Gif-sur-Yvette Cedex, France*

ABSTRACT

After reviewing briefly the role played by singular vectors in the representation theory of Lie algebras, we recall the main properties of singular vectors in Virasoro Verma modules. Then we describe procedures based on fusion to calculate these singular vectors explicitly.

1. INTRODUCTION

The fundamental role played by the Virasoro algebra in 2-d conformal field theories^[1] has motivated much work, and the structure of its irreducible highest weight representations is now fairly well understood. Among these, those which are unitarisable are of utmost importance in statistical mechanics. The key concept for testing irreducibility and unitarity (though this is somewhat more involved) is that of singular vectors, playing in a sense the role of the top spin component in the case of angular momentum. Although a lot is known about these singular vectors (the fundamental results were conjectured by Kac^[2] and later proved by Feigin and Fuchs^[3,4], initiating a long series of work on free field representations^[5]) no general explicit form has been proposed. We will describe in the following pages an efficient method to compute them, establishing along the way an intimate connection between singular vectors, operator products and fusion rules. This work was done in collaboration with P. Di Francesco, C. Itzykson and J.B. Zuber. We shall present our results here using an operator formalism and concentrating on the algebraic aspects, which in turn lead to a fairly pedagogical introduction to several key constructions in conformal field theory. The reader interested in the more geometrical aspects and in particular the link between singular vectors and W -algebras is referred to the original papers [6,7].

In section 2 we describe the fundamental concepts, quoting the theorems of Feigin and Fuchs. In section 3 we recall some elementary facts concerning operator product

[†] Laboratoire de la Direction des Sciences de la Matière du Commissariat à l'Energie Atomique

expansion. In section 4 we construct a particular class of singular vectors, using them for the general construction in section 5. The last section is devoted to some comments.

2. SINGULAR VECTORS AND REPRESENTATION THEORY

2.1. Angular momentum as an illustrative example

Let us consider the Lie algebra A_1 generated by J_- , J_0 , J_+ with commutation relations

$$[J_0, J_{\pm}] = \pm J_{\pm} \quad [J_+, J_-] = 2J_0$$

Usually physicists construct the representations of this algebra by demanding that J_+ and J_- should be mutually adjoint and looking for a positive definite scalar product on the Hilbert space carrying the representation. On the other hand, mathematicians look for finite dimensional irreducible representations, and prove that J_0 is semi-simple (i.e. can be diagonalized). Due to a classical theorem, non zero eigenvectors of J_0 with distinct eigenvalues are linearly independent. So if $|v\rangle$ is an eigenvector of J_0 with lowest eigenvalue $-j$, the vectors $J_+^k|v\rangle$ (if non zero) are eigenvectors with eigenvalue $-j+k$, and are linearly independent. Hence there is a minimal k , call it $k_0 > 0$, such that $J_+^{k_0}|v\rangle = 0$, and using the commutation relation

$$[J_+^k, J_-] = k J_+^{k-1} (2J_0 + k - 1)$$

we get

$$[J_+^{k_0}, J_-]|v\rangle = k_0(k_0 - 2j - 1)J_+^{k_0-1}|v\rangle$$

But the left hand side is zero ($J_+^{k_0}|v\rangle = 0$ by hypothesis and $J_-|v\rangle = 0$ because $|v\rangle$ has lowest J_0 eigenvalue) and $J_+^{k_0-1}|v\rangle$ is non zero. Hence $k_0 = 2j + 1$, and j has to be an integer or half integer. Although the calculations are very similar for the physicist and for the mathematician, there is a big difference in philosophy. The mathematician has the following reinterpretation: for complex λ consider a representation V_λ of A_1 with the two properties: (i) it is cyclic i.e. generated by one vector $|v\rangle$ and this vector $|v\rangle$ is annihilated by J_- and is an eigenvector of J_0 with eigenvalue λ , (ii) it has the following universal property: any other representation satisfying (i) is isomorphic to a quotient of V_λ by a subrepresentation. V_λ is called a Verma module. Its universal property makes it unique if it exists, and it is easy to show that a formal vector space with basis $\{J_+^k|v\rangle, k \in \mathbb{Z}_+\}$ can be endowed with a representation structure defining V_λ . Then it is possible to prove that V_λ is irreducible except when $-\lambda$ is an integer or half integer. If $-\lambda = j$ with $2j$ integer then the state $J_+^{2j+1}|v\rangle$ is annihilated by J_- (just like $|v\rangle$), and it generates a new Verma module V_j , the quotient of $V_{-\lambda}$ by V_j being isomorphic to the finite dimensional representation we constructed before. Usually $|v\rangle$ is called a highest weight state (because it has the lowest J_0 eigenvalue; once more physics and mathematics have opposite conventions) and

$J_+^{2j+1}|v\rangle$ is called a singular vector, because it is set to zero in the quotient. Finally, let us remark that V_λ never has a positive definite scalar product making J_+ and J_- adjoint of each other.

The same kind of techniques can be applied to any complex semisimple Lie algebra and can be used to compute characters for instance.

2.2. The Virasoro algebra

The Virasoro algebra is a central extension of the algebra of vectors fields on the circle. Explicitly it is generated by L_n , $n \in \mathbb{Z}$ and c , a central element (which in the following assumes a numerical value), with commutation relations:

$$[L_m, L_n] = (m - n) L_{m+n} + \frac{c}{12} (m^3 - m) \delta_{m+n}$$

Just as in the case of A_1 , this algebra is naturally graded (by L_0 instead of J_0) and L_{-1} , L_0 , L_1 have the commutation relations of a non compact form of A_1 . It turns out that (for independant reasons, not yet completely understood), A_1 plays a deep role in the representation theory of the Virasoro algebra.

Just as before we consider a Verma module $V_{c,h}$ as a representation of the Virasoro algebra with the following properties: (i) $V_{c,h}$ contains a highest weight state $|h\rangle$, annihilated the L_n 's with $n > 0$, $L_0|v\rangle = h|h\rangle$, and $|h\rangle$ generates the whole representation by action of the generators (i.e. $V_{c,h}$ is cyclic), (ii) any representation satisfying (i) is a quotient of $V_{c,h}$. One can verify that the formal vector space with basis $\{L_{-n_1}L_{-n_2}\dots L_{-n_k}|v\rangle, 1 \leq n_1 \leq n_2 \leq \dots \leq n_k, k \geq 0\}$ can be endowed with the appropriate representation structure, showing the existence of $V_{c,h}$. Once again property (ii) then shows uniqueness up to isomorphism. In a physical context, the fact that L_0 is bounded below is natural since L_0 is related to the energy.

Singular vectors appear when examining the irreducibility of $V_{c,h}$, which is naturally graded by the eigenvalues of L_0 : $V_{c,h} = \bigoplus_{k \geq 0} V_k$ where L_0 restricted to V_k acts as multiplication by $h + k$. A standard argument shows that any submodule M of $V_{c,h}$ is also graded: Any vector in $V_{c,h}$, hence in M , is a finite sum of eigenvectors of L_0 . If $e = \sum_k e_k$, ($e_k \in V_k$) belongs to M then so does $L_0^\ell e = \sum_k (h + k)^\ell e_k$. When ℓ runs over the non-negative integers we obtain linear combinations of the e_k 's belonging to M . The number of non vanishing e_k 's is finite, and by solving a finite linear system with non vanishing determinant (a Vandermonde determinant with distinct arguments) one proves that all the e_k 's belong to M , showing that $M = \bigoplus_k M \cap V_k$. If k_0 is the smallest k such that $M \cap V_k \neq \{0\}$ then $M \cap V_{k_0}$ is annihilated by all the L_n 's with $n > 0$. Hence, either $k_0 = 0$ and from cyclicity $M = V_{c,h}$ or $k_0 \neq 0$ and M contains a singular vector. By acting with the L_n 's with $n < 0$ on this vector we build a representation of the Virasoro algebra isomorphic to $V_{c,h+k_0}$.

2.3. Fundamental results

Beside proving the conjectured Kac determinantal formula, Feigin and Fuchs exhibited numerous properties of singular vectors. After fixing the notations and giving the main theorem, we shall list some of these.

We introduce a complex parameter θ and two (a priori complex) numbers j and j' and set

$$\begin{aligned} t &= \theta^2, \\ c(\theta) &= 1 + 6(\theta + \theta^{-1})^2, \\ h_{j',j}(\theta) &= -(j\theta + j'\theta^{-1})((j+1)\theta + (j'+1)\theta^{-1}) \end{aligned}$$

Then we have the following

Theorem (Kac, Feigin and Fuchs): $V_{c,h}$ contains a singular vector at level k if and only if, for a certain complex number θ and a certain pair (j', j) of non-negative integers or half integers,

$$c \equiv c(\theta), \quad h \equiv h_{j',j}(\theta), \quad k \equiv (2j+1)(2j'+1)$$

It is clear that the parametrization with θ , j' and j is cleverly chosen and already suggests a connection with representations of angular momentum via j' and j . As a first consequence of the theorem we see that for generic θ (for instance such that $t = \theta^2$ is not quadratic over Q) a Verma module contains at most one singular vector (up to normalisation). In general the theorem fully describes all the Verma submodules included in a given $V_{c,h}$, and the way the mutually intersect.

Our goal will be to give explicit expressions for the singular vectors or at least efficient algorithms for their computation. For fixed θ we shall call $\phi_{j',j}$ the singular vector corresponding to $h_{j',j}$. Using the basis for $V_{c,h}$ mentioned above we can associate to $\phi_{j',j}$ a polynomial in the L_{-n} 's. We shall use the notation $\phi_{j',j}(L)$. Feigin and Fuchs proved that:

(i) In $\phi_{j',j}(L)$ the coefficient of $L_{-1}^{(2j+1)(2j'+1)}$ is non vanishing. In the following we put this coefficient equal to one. Then $\phi_{j',j}(L)$ is a finite Laurent series in the variable t . Feigin and Fuchs conjecture that when t goes to $-t$ and L_{-n} to $(-1)^{n-1}L_{-n}$, (hence c goes to $26-c$) $\phi_{j',j}$ remains invariant.

(ii) The L_{-n} , $n \geq 1$, span a Lie algebra without central term, and the L_{-n} , $n \geq n_0$, span an ideal in this Lie algebra, call it I_{n_0} . We remark that (for any c and h) $V_{c,h}$ is isomorphic, as an I_1 representation to the universal enveloping algebra of I_1 , and the value of $\phi_{j',j}$ modulo I_n is a well defined concept. The first part of property (i) gives the case $n = 2$. Feigin and Fuchs also describe the case $n = 3$. Modulo I_3 , L_{-1} and L_{-2} commute and

$$\phi_{j,j'}^2 = \prod_{\substack{-j \leq M \leq j \\ -j' \leq M' \leq j'}} \left(L_{-1}^2 + 4(M\theta + M'\theta^{-1})^2 L_{-2} \right)$$

which in turn is a perfect square.

(iii) For the same reasons we can specialize $\phi_{j,j'}$ to any representation of I_1 , for instance $\ell_{-n}\psi_m = (\alpha(n-1) - \beta - m)\psi_{m-n}$. This is an action of I_1 by differential operators on functions of the form $z^\beta f(z)$ where $f(z)$ is a formal Laurent series that we can truncate at some maximal degree if we want. We shall see that this representation plays a crucial role in the discussion. Feigin and Fuchs give the value of $\phi_{j,j'}$ restricted to this representation as a fairly complicated product over M and M' of polynomials of degree 2 in β and 1 in α , once more with a typical “angular momentum” flavor.

Benoit and Saint-Aubin^[8] gave explicit formulae for the singular vectors in case j' or j vanishes. We shall recover these with a new interpretation in section 4. Apart from this very little seems to be known in general. The calculation of the singular vectors $\phi_{j',j}$ by brute force for small j' , j is very tedious.

3. OPERATOR PRODUCT EXPANSIONS AND FUSION

To avoid mathematical complications, we shall imagine that we are looking at the holomorphic part of a “big” conformal field theory, that is we consider a vector space E carrying a representation of the Virasoro algebra with given central charge. We assume that E splits as a direct product of irreducible highest weight representation of the Virasoro algebra, one of the factors being the vacuum sector E_0 with highest weight state $|\Omega\rangle$ of weight 0, annihilated by the L_n 's with $n \geq -1$, i.e. projective invariant.

We call a family of operators $\varphi(z)$ on E a chiral primary field of weight h if

$$[L_n, \varphi(z)] = \left((n+1)h z^n + z^{n+1} \frac{d}{dz} \right) \varphi(z), \quad n \in \mathbb{Z} \quad (1)$$

It is clear that $\varphi(0)|\Omega\rangle$ is a highest weight state of weight h . The main simplification in our approach is that we can assume that $\varphi(z)$ acts on the whole of E , so that we can make products of chiral primary fields and then decompose them as linear operators between the different highest weight representations building E . In a more rigorous treatment we should start with representations of the Virasoro algebra, construct the deformed tensor product of Moore and Seiberg^[9], then the chiral primary fields would be seen as intertwiners, the intertwining property basically having the same meaning as relation (1). We would then have to work hard to make these vertex operators act as linear operators that we can compose. We shall treat the primary fields, just in the way we manipulate the spin operator for the Ising model, as operators acting directly on the full Hilbert space.

3.1. Covariance in operator product expansions

Let φ_0 and φ_1 be two chiral primary fields of respective weights h_0 and h_1 . As usual in conformal field theory we expect that the product $\varphi_0(z_0)\varphi_1(z_1)$ decomposes as a linear

combination of local scaling operators. To be more precise if we set $z_0 = z + \xi(\lambda + \frac{1}{2})$, $z_1 = z + \xi(\lambda - \frac{1}{2})$ where λ is some fixed constant we expect that

$$\varphi_0(z_0)\varphi_1(z_1) = \sum_{\Delta} \xi^{\Delta-h_0-h_1} \mathcal{O}_{\Delta}(z) \quad (2)$$

We assume that the Δ 's appearing in this sum are bounded below. In the sequel we shall often take $\lambda = 1/2$, but any λ will do. The two sides of Eq.(2) should have the same conformal properties, that is the same commutators with the L_n 's. On the left hand side, these act as differential operators, and we have

$$\begin{aligned} & \sum \xi^{\Delta-h_0-h_1} [L_n, \mathcal{O}_{\Delta}(z)] \\ &= \left(h_0(n+1)z_0^n + z_0^{n+1} \frac{\partial}{\partial z_0} + h_1(n+1)z_1^n + z_1^{n+1} \frac{\partial}{\partial z_1} \right) \sum_{\Delta} \xi^{\Delta-h_0-h_1} \mathcal{O}_{\Delta}(z) \end{aligned} \quad (3)$$

On the left hand side of (3) we expand z_0 and z_1 in powers of ξ and we see that $[L_n, \mathcal{O}_{\Gamma}(z)]$ is given by differential operators acting on the $\mathcal{O}_{\Delta}$'s with $\Gamma - \Delta \in \mathbb{Z}$. For instance looking at the case $n = -1$ we observe that $[L_{-1}, \mathcal{O}_{\Delta}(z)] = \partial \mathcal{O}_{\Delta}(z)$, so that in general L_{-1} generates translations. Moreover if there is no Δ such that $\Gamma - \Delta$ is a positive integer we see that $\mathcal{O}_{\Gamma}$ has to be a chiral primary field of weight Γ (in fact it is easy to see that the differential operator on the left hand side of (3) never decreases the exponent of ξ , and then to check that $[L_n, \mathcal{O}_{\Gamma}(z)]$ has the form expected for a primary field of weight Γ). The same reasoning shows that for any Γ

$$[L_n, \mathcal{O}_{\Gamma}(z)] - \left(\Gamma(n+1)z^n + z^{n+1} \frac{\partial}{\partial z} \right) \mathcal{O}_{\Gamma}(z)$$

depends only on $\mathcal{O}_{\Delta}$'s with $\Gamma - \Delta$ a positive integer. So, in the expansion we can concentrate on sectors starting with a chiral primary field, getting for a given weight h :

$$\begin{aligned} & \sum_{k=0}^{\infty} \xi^{h-h_0-h_1+k} [L_n, \mathcal{O}_{h+k}(z)] \\ &= \left(h_0(n+1)z_0^n + z_0^{n+1} \frac{\partial}{\partial z_0} + h_1(n+1)z_1^n + z_1^{n+1} \frac{\partial}{\partial z_1} \right) \sum_{k=0}^{\infty} \xi^{h-h_0-h+k} \mathcal{O}_{h+k}(z) \end{aligned} \quad (3')$$

These equations for the $\mathcal{O}$'s are linear. If there are two distinct solutions starting with the same $\mathcal{O}_h$ then they begin to differ at a certain level k and the difference at level k has to be a primary of weight $h+k$ by the preceding argument. If there is no primary field of weight h' with $h' - h$ a positive integer in the theory we are considering then Eq.(3') can have at most one (up to a factor) non vanishing solution.

To be more specific we can evaluate Eq.(3') on the vacuum $|\Omega\rangle$ of the theory and look for solutions such that $\mathcal{O}_{h+k}(z)|\Omega\rangle$ belongs to a representation of the Virasoro algebra. For

the sake of simplicity we take $\lambda = \frac{1}{2}$ and, restricting to non negative n 's we can set $z = 0$. We denote by $|k\rangle$ the state $\mathcal{O}_{h+k}(0)|\Omega\rangle$. If we recall that L_1, L_2 generate by commutation all the L_n 's with n positive, we obtain following equations

$$\begin{aligned} L_0|0\rangle &= L_2|0\rangle = 0 & L_1|1\rangle &= 0 \\ L_0|k\rangle &= (h+k)|k\rangle \\ L_1|k\rangle &= (h+h_0-h_1+k-1)|k-1\rangle & k \geq 1 \\ L_2|k\rangle &= (h+2h_0-h_1+k-2)|k-2\rangle & k \geq 2 \end{aligned} \quad (4)$$

Of course these equations say that $|0\rangle$ is a highest weight state of weight h , and we can look for solutions in the cyclic module M that it generates. Given $|k-2\rangle, |k-1\rangle$ this gives a linear systems for $|k\rangle$. If $V_{c,h}$ is irreducible then $M \cong V_{c,h}$, L_1 and L_2 have no common vector in their kernels except $|0\rangle$, and there is at most one solution. On the other hand if $V_{c,h}$ is reducible L_1 and L_2 annihilate some vector in $V_{c,h}$. If the class of this vector is non zero in M some arbitrariness appears. If it is zero then this puts constraints on the possible values of h_0 and h_1 . As a trivial example consider the case $n = 0$. Then $L_{-1}|0\rangle$ is a singular vector in $V_{c,0}$. If its class is zero in M (for instance if M is the vacuum sector), then $|1\rangle = 0$ and $L_1|1\rangle = (h_0 - h_1)|0\rangle$ so h_0 and h_1 must be equal. Other less trivial examples are to be found in [7]. At level n the Verma module $V_{c,h}$ contains $P(n)$ states and looking for a singular vector amounts to solving a $P(n) \times P(n)$ linear system. Happily there is another procedure which we now describe.

4. SINGULAR VECTORS OF TYPE $(j', 0)$ OR $(0, j)$

It is very easy to check that

$$\left(L_{-1}^2 + \frac{1}{t}L_{-2}\right)\varphi_1(0)|\Omega\rangle$$

is a singular vector corresponding to $j' = 1/2$ and $j = 0$ if $\varphi_1(z)$ is the chiral primary field associated to the highest weight state of weight $h_1 = -\left(\frac{1}{2} + \frac{3}{4t}\right) = h_{1/2,0}$. This means that in the irreducible module we have $(L_{-1}^2 + \frac{1}{t}L_{-2})\varphi_1(0)|\Omega\rangle = 0$. If $\varphi_0(z)$ is a primary field of weight h_0 we have

$$\varphi_0(z)\left(L_{-1}^2 + \frac{1}{t}L_{-2}\right)\varphi_1(0)|\Omega\rangle = 0$$

that is, using (1)

$$\left(\left(L_{-1} - \frac{\partial}{\partial z}\right)^2 + \frac{1}{t}\left(L_{-2} + \frac{h_0}{z^2} - \frac{1}{z}\frac{\partial}{\partial z}\right)\right)\varphi_0(z)\varphi_1(0)|\Omega\rangle = 0$$

We change z into $-z$ and conjugate with the operator $e^z L_{-1}$ (making a translation of argument z) to get

$$e^z L_{-1}\left(\left(L_{-1} + \frac{\partial}{\partial z}\right)^2 + \frac{1}{t}\left(L_{-2} + \frac{h_0}{z^2} - \frac{1}{z}\frac{\partial}{\partial z}\right)\right)e^{-z} L_{-1}e^z L_{-1}\varphi_0(-z)\varphi_1(0)|\Omega\rangle = 0$$

If we take the operator product expansion in a given sector of weight h , $e^{z L_{-1}} \varphi_0(-z) \varphi_1(0) e^{-z L_{-1}}$ is simply $\varphi_1(z) \varphi_0(0)$ up to a phase. We write $\varphi_1(z) \varphi_0(0) = \sum_{k=0}^{\infty} z^{h-h_0-h_1+k} \mathcal{O}_{h+k}(0)$ and we obtain

$$\left(\frac{\partial^2}{\partial z^2} + \frac{h_0}{tz^2} - \frac{1}{tz} \frac{\partial}{\partial z} + \frac{1}{t} \sum_{n \geq 1} L_{-n} z^{n-2} \right) \sum_{k \geq 0} z^{h-h_0-h_1+k} |k\rangle = 0 \quad (5)$$

This gives recursion relations among the $|k\rangle$'s of the form

$$f(k)|k\rangle = -\frac{1}{t} \sum_{\substack{m+n=k \\ n \geq 1}} L_{-n}|m\rangle$$

where

$$f(k) \equiv \left[(h-h_0-h_1+k)(h-h_0-h_1+k-1) + \frac{h_0}{t} - \frac{h-h_0-h_1+k}{t} \right].$$

For $k=0$ this gives

$$\left[(h-h_0-h_1)(h-h_0-h_1-1) + \frac{h_0}{t} - \frac{h-h_0-h_1}{t} \right] |0\rangle = 0$$

If $|0\rangle = 0$ then all $|k\rangle$'s are 0 so this gives a selection rule: for $k=0$, f has to be zero for fusion to be possible. We see that the presence of the singular vector for $h_{1/2,0}$ makes the fusion much easier to calculate. If $f(0) = 0$ we can write in general $f(k) = k(k-2j-1)$ for a certain j . Then we find $h = h_{0,j}$, $h_0 = h_{-1/2,j}$. Hence we are calculating the fusion $(\frac{1}{2}, 0) \times (-\frac{1}{2}, j) \rightarrow (0, j)$. We note that whatever j is, $(-\frac{1}{2}, j)$ has no singular vector associated to it. A small miracle happens when $2j+1$ happens to be a positive integer. Then our recursion relations determine $|1\rangle, \dots, |2j\rangle$ in terms of $|0\rangle$ and the relation at $|2j+1\rangle$ degenerates to give

$$\sum_{\substack{m+n=2j+1 \\ n \geq 1}} L_{-n}|m\rangle = 0.$$

The meaning is as follows: fusion has "transported" the singular vector at level 2, $(\frac{1}{2}, 0)$, to give the singular vector at level $2j+1$, $(0, j)$. We could make the same manoeuvre to get $(j', 0)$ from $(0, \frac{1}{2})$. It is of course possible to check directly that if for $h_{0,j}$, we define the vectors $|1\rangle, \dots, |2j\rangle$ using our recursion relations then

$$\sum_{\substack{m+n=2j+1 \\ n \geq 1}} L_{-n}|m\rangle = 0$$

is annihilated by the L_n 's $n > 0$.

All this is extremely simple and efficient. Properties (i) and (ii) of Feigin and Fuchs are (of course!) true for our expressions. There is at this point an unexpected and beautiful connection with the theory of classical W -algebras. The interested reader should consult [7].

Our expressions are just the Benoit Saint-Aubin formulae, recasted in matrix form (or recursive form) and interpreted in terms of fusion with the consequence that the descent equations (4) are valid. Moreover the representation associated to angular momentum j naturally appears. To see this let us choose as a basis for the $2j + 1$ representation of A_1 the elements $J_+^k v$ for $k = 0$ to $2j$ (where v is the highest weight state in the A_1 representation). Then $k(2j + 1 - k)$ is the corresponding matrix element of J_- and if we denote by $|(0, j)\rangle$ the column vector $\begin{pmatrix} |2j\rangle \\ \vdots \\ |0\rangle \end{pmatrix}$ we can rewrite our result as

$$\left(-J_- + \frac{1}{t} \sum_{n=0}^{2j} J_+^n L_{-n-1}\right) |(0, j)\rangle = 0, \quad (6)$$

Of course there is a similar equation for the $(j', 0)$ case, namely

$$\left(-J_- + t \sum_{n=0}^{2j'} J_+^n L_{-n-1}\right) |(j', 0)\rangle = 0 \quad (6')$$

5. GENERAL SINGULAR VECTORS

We can now play the same game to get expressions for the (j, j') singular vector using fusion. As we shall see this procedure has several drawbacks. However, it is interesting, gives efficient algorithms and illustrates the power of elementary conformal field theory.

We start from Eq.(6) and multiply on the left by a chiral primary field of weight h_0 . After commutation with the L_n 's we get

$$\left(-J_- - \frac{1}{t} \frac{z}{z - J_+} + \frac{h_0}{t} \frac{J_+}{(z - J_+)^2} + \frac{1}{t} \sum_{n=0}^{2j} L_{-n-1} J_+^n\right) \varphi_0(z) |(0, j)\rangle = 0$$

We are going to concentrate, in the operator product expansion of $\varphi_0(z)$ with the components of $|(0, j)\rangle$, on the sector of highest weight h . This sector is stable under the action of the Virasoro generators, so we can, as in the previous section, change z into $-z$ and conjugate by $e^{zL_{-1}}$. We use the fact that

$$e^{zL_{-1}} \left(\sum_{n=0}^{2j} L_{-n-1} J_+^n \right) e^{-zL_{-1}} = \frac{1}{t} \left(L_{-1} + \sum_{n=2}^{\infty} L_{-n} J_+ (z + J_+)^{n-2} \right).$$

Then

$$\left(-J_- + \frac{1}{t} \frac{z}{z+J_+} \partial + \frac{h_0}{t} \frac{J_+}{(z+J_+)^2} + \frac{1}{t} \sum_{n=1}^{\infty} L_{-n} J_+ (z+J_+)^{n-2} \right) e^z L_{-1} \varphi(-z) |(0, j)\rangle = 0$$

When we expand in the h sector we expect that the power of z will be $L_0 - h_0 - h_1 - J_0 - j$ (where h_1 is now $h_{0,j}$), and so we end up with the operator

$$\left(-J_- + \frac{1}{t} \frac{1}{(1+J_+)} (L_0 - h_0 - h_1 - J_0 - j) + \frac{h_0}{t} \frac{J_+}{(1+J_+)^2} + \frac{1}{t} \sum_{n=1}^{\infty} L_{-n} J_+ (1+J_+)^{n-2} \right)$$

acting on some $2j$ component vector (with components in the h sector) to give 0. For this equation to have non zero solutions, the operator acting at level 0 (i.e. $L_0 - h = 0$) has to have a kernel. So we find once more a selection rule saying that h_0 and h have to satisfy a relation, namely

$$\text{Det} \left(-J_- + \frac{1}{t} \frac{1}{(1+J_+)} (h - h_0 - h_1 - J_0 - j) + \frac{h_0}{t} \frac{J_+}{(1+J_+)^2} \right) = 0$$

The construction shows that this determinant is given by property (iii) from Feigin and Fuchs with α and β simply related to the conformal weights h, h_0, h_1 .

To calculate the determinant directly we use the following trick: by conjugation with $(1+J_+)^{\gamma}$ with appropriate γ we get rid of the term containing $\frac{J_+}{(1+J_+)^2}$. We conjugate the whole operator, multiply by $1+J_+$ and obtain finally:

$$\left(-J_- + \frac{1}{t} (L_0 - h_0 + t J_0^2 - (1+2\gamma t + t) J_0) + \frac{1}{t} \sum_{n=1}^{\infty} L_{-n} J_+ (1+J_+)^{n-1} \right)$$

where γ is such that $h_0 = -\gamma - t \gamma(\gamma+1)$, i.e. $h_0 \equiv h_{0,\gamma}$. The selection rule is now

$$\prod_{-j \leq m \leq j} (h - h_0 + t m^2 - (1+2\gamma t + t)m) = 0$$

Instead of determining the vector at level k in terms of vectors at a lower level, this construction will give relations (they could be identities) among them if at level k (i.e. $L_0 - h = k$) the matrix

$$M_k = -J_- + \frac{1}{t} (h + k - h_0 + t J_0^2 - (1+2\gamma t + t) J_0) \quad (7)$$

is singular, i.e. if

$$\prod_{-j \leq m \leq j} (h - h_0 + t m^2 - (1+2\gamma t + t)m + k) = 0$$

For instance if we choose $h_0 = h_{j', j-1/2}$ we can take $\gamma = \frac{j'}{t} - \frac{1}{2}$ and our operator becomes

$$-J_- + \frac{1}{t}(L_0 - h - (2j' + 1)(J_0 + j)) + (J_0 + j)(J_0 - j) + \frac{1}{t} \sum_{n=1}^{\infty} L_{-n} J_+ (1 + J_+)^n$$

For generic (irrational) t the matrix is singular if and only if $L_0 = h$ (then an eigenvector of J_0 with eigenvalue $-j$ is in the kernel) or $L_0 = h + 2j(2j' + 1)$ (then an eigenvector of J_0 with eigenvalue j cannot be in the image, meaning that if $|1\rangle, \dots, |2j(2j' + 1) - 1\rangle$ are the vectors determined by solving the recursion relations starting from $|0\rangle$ the top ($J_0 = j$) component of

$$\sum_{\substack{m+n=k \\ n \geq 1}} L_{-n} J_+ (1 + J_+)^n |m\rangle$$

has to vanish, giving the expected singular vector corresponding to the fusion $(0, j) \times (j', -\frac{1}{2}) \rightarrow (j', j - \frac{1}{2})$. So we finally obtain explicit algorithms to compute the singular vectors. They are efficient because the matrix M_k we have to invert at each intermediate step is lower triangular, the vector we start with is simple (eigenvector of J_0 with eigenvalue $-j$) and the singular vector is just the top spin component ($J_0 = j$) at the last step.

We have choosen here the simplest fusion that gives (j', j) namely $(0, j + \frac{1}{2}) \times (j', -\frac{1}{2}) \rightarrow (j', j)$. Among other possibilities we could have choosen any angular momentum $J > j$ and any non negative integer p such that $p + 2j + 1 \leq 2J$ to obtain the fusion $(0, J) \times (j', p + j - J) \rightarrow (j', j)$.

6. COMMENTS AND CONCLUSIONS

The last section exhibited matrix forms for the expressions of singular vectors. However let us mention another, closely related, natural form for these vectors. Equation (5) was a differential equation with operator coefficients. Our requirements in section 4, leading to fusion rules, determined the exponent of the Froebenius series (here with only one term) solution of the purely differential part. The recursion relations simply determine the successive (vector) coefficients of an analytic function solution of

$$\left(\frac{\partial^2}{\partial u^2} - 2j \frac{1}{u} \frac{\partial}{\partial u} + \frac{1}{t} \sum_{m \geq 1} L_{-m} u^{m-2} \right) f(u) = 0$$

were we recognize $T_-(u) = \sum_{m \geq 1} L_{-m} u^{m-2}$, the negative part of the stress tensor.

In fact we simply deal with a Verma representation of A_1 where J_+ is multiplication by u , J_0 is the operator $u \frac{\partial}{\partial u} - j$ and J_- the operator $u \frac{\partial^2}{\partial u^2} - 2j \frac{\partial}{\partial u}$. We can do the same for the general singular vector, ending up this time with general second order fuchsian equation perturbed by the negative part of the stress energy tensor. The former $\frac{J_+}{(1+J_+)^2}$ is replaced

by a double pole in the differential equation, and the trick with γ we used to eliminate it, is simply the standard change of function used to put the fuchsian equation in hypergeometric form. The fusion rules now just state that we are looking (in the unperturbed problem) for a solution in the form of a Frobenius series without logarithmic singularities at the origin. This approach does not seem well suited for explicit calculations because perturbations of the hypergeometric equation are not easy to handle, but the physical interpretation is very clear and allows to make contact with the differential equations satisfied by the insertions of primary fields in correlations functions.

We believe that a good reinterpretation of our results should give an alternative (elementary) proof of Kac's determinantal formula.

However in the calculation of the general singular vector, we have broken the symmetry between j' and j , hence have no natural angular momentum appearing anymore in our expressions, and all the properties of Feigin and Fuchs are very difficult to check. We hope to return to this problem in the future.

7. REFERENCES

- [1] A. Belavin, A. Polyakov and A. Zamolodchikov, Nucl. Phys. **B241** (1984), 333-380.
- [2] V. Kac, Lect. Notes in Math. **94** (1979) 441-445.
- [3] B.L. Feigin and D.B. Fuchs, Funct. Anal. & Appl. **17** (1982), 114-126 [Funk. Anal. Priloz. **16**, No.2 (1982), 47-63].
- [4] B.L. Feigin and D.B. Fuchs, Lect. Notes in Math. **1060** (1984) 230-245.
- [5] G. Felder, Nucl. Phys. **B237** (1989) 215-236, err. *ibid* **B234** (1989) 548.
- [6] P. Di Francesco, C. Itzykson and J.-B. Zuber, *Classical W-Algebras*, Saclay preprint SPhT/90-149, to be published in Comm. Math. Phys.;
C. Itzykson, *W-Geometry*, to appear in the proceedings of the 1990 Cargèse Workshop, and further references therein.
- [7] M. Bauer, P. Di Francesco, C. Itzykson and J.-B. Zuber, *Covariant Differential Equations and Singular Vectors in Virasoro Representations*, Saclay preprint SPhT/91-030, to be published in Nucl. Phys. B.
- [8] L. Benoit and Y. Saint-Aubin, Phys. Lett. **215B** (1988) 517-522.
- [9] G. Moore and N. Seiberg, *Lectures on RCFT*, Lectures at the 1989 Trieste Spring School.

DOUBLE SCALING LIMIT IN $O(N)$ VECTOR MODELS

P. Di Vecchia

NORDITA, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

Abstract

We discuss some recently found results in the $O(N)$ vector models in D dimensions within the framework of the large N expansion. In particular we show that the Green's functions of those models present a singularity for a certain value of the coupling constant in every order of the $1/N$ expansion. It is possible to perform a double scaling limit exactly as in the matrix models obtaining a well defined scaling theory. The advantage of these models with respect to the matrix models is that this procedure can be extended to dimensions higher than just 0 and 1. The critical point is characterized by the fact that a composite field become massless.

Recently a double scaling limit in matrix models in zero and one space-time dimension has been performed¹⁾. In this limit the dimension of the matrix $N \rightarrow \infty$ and the coupling constant g approaches a critical value g_c in such a way that the scaling variable $t = N(g - g_c)^\alpha$ is kept fixed. α is a suitable constant. This limit corresponds to the nonperturbative continuum limit of the discretized string²⁾.

Such critical behaviour is not unique of the matrix models, but has been also obtained in the $O(N)$ vector models both for $D = 0^{3,4,5)}$ and for $D \geq 1^{6,7)}$ where the large N expansion can also be performed in a very straightforward way.

In this talk I want to summarize the results obtained in collaboration with Mitsuhiro Kato and Nobuyoshi Ohta in the $O(N)$ vector models and already published^{5,7)}.

The simplest example of a vector model is described by the action

$$S = \int d^D x \left[\frac{1}{2} \partial_\mu \vec{n} \partial_\mu \vec{n} + \frac{1}{2} \beta^2 \vec{n}^2 + g(\vec{n}^2)^2 \right] \quad (1)$$

In order to perform the large N expansion it is convenient to make the action quadratic in the elementary field $\vec{n}$ by introducing an auxiliary field σ . In terms of this field the action in eq. (1) becomes

$$S = \int d^D x \left[\frac{1}{2} (1 + K_n) \partial_\mu \vec{n} \partial_\mu \vec{n} + \frac{1}{2} \beta^2 (1 + K_m) \vec{n}^2 - \frac{1}{2} i \sigma \vec{n}^2 + \frac{N}{16 f \mu^{2\epsilon} (1 + K_f)} \sigma^2 \right] \quad (2)$$

where we have introduced the renormalization constants that are necessary to eliminate the ultraviolet divergences for $D \geq 2$ and $f \equiv Ng$. f is kept fixed in the large N expansion.

The integral over the N -component field $\vec{n}$ can then be performed arriving at the following expression for the partition function

$$Z(f) = \int D\sigma e^{-S_{eff}} \quad (3)$$

where

$$S_{eff} = \int d^D x \frac{1}{16f(1+K_f)} \sigma^2 + \frac{1}{2} Tr \log [-(1+K_n)\partial^2 + \beta^2(1+K_m) - i\sigma] \quad (4)$$

Expanding the effective action around the saddle point

$$\sigma(x) = \sigma_0 + \alpha(x) \quad (5)$$

satisfying the saddle point equation

$$\frac{\sigma_0}{4f(1+K_f)} = i \int \frac{d^D p}{(2\pi)^D} \frac{1}{(1+K_n)p^2 + m^2} = \frac{i}{(1+K_n)} \frac{\Gamma(1-D/2)}{(4\pi)^{D/2}(m_R^2)^{1-D/2}} \quad (6)$$

we get

$$S_{eff} = S_{eff}^{(0)} + \frac{1}{2} \int d^D x \alpha(x) \int d^D y \alpha(y) \int \frac{d^D k}{(2\pi)^D} e^{ik \cdot (x-y)} \Gamma(k) + \\ - \frac{1}{2} \sum_{n=3}^{\infty} \frac{(i)^n}{n} \int \prod_{i=1}^n [d^D x_i] \prod_{i=1}^n \left[\frac{d^D p_i}{(2\pi)^D} e^{ip_i \cdot x_i} \right] (2\pi)^D \delta(\sum_{i=1}^n p_i) \Delta(p_1, p_2, \dots, p_n) \quad (7)$$

where

$$S_{eff}^{(0)} = \frac{1}{16f(1+K_f)} \int d^D x \sigma_0^2 + \frac{1}{2} Tr \log [-(1+K_n)\partial^2 + m^2] \quad (8)$$

and

$$m^2 = \beta^2(1+K_m) - i\sigma_0 \quad m_R^2 \equiv m^2/Z_n \quad Z_n \equiv 1+K_n \quad (9)$$

In addition we have

$$\Gamma(k) = \frac{1}{8f(1+K_f)} + \frac{1}{2} \int \frac{d^D p}{(2\pi)^D} \frac{1}{[(1+K_n)p^2 + m^2][(1+K_n)(p-k)^2 + m^2]} \quad (10)$$

and

$$\Delta(p_1, p_2, \dots, p_n) = \int \frac{d^D k}{(2\pi)^D} \frac{1}{[Z_n k^2 + m^2][Z_n(k+p_1)^2 + m^2] \cdots [Z_n(k+p_1+p_2+\cdots+p_n)^2 + m^2]} \quad (11)$$

After renormalization one can see that there is a critical value of the coupling constant in every order of the $1/N$ expansion, that can be obtained either from the potential

$$V(r) = \frac{1}{2}\beta^2 r^2 + fr^4 + \begin{cases} -\log r & D=0 \\ 1/(8r^2) & D=1 \\ \frac{\beta^2}{8\pi}e^{-4\pi r^2} & D=2 \\ -\frac{8\pi^2}{3}r^6 & D=3 \end{cases} \quad (12)$$

by imposing the two conditions

$$V'(r) = V''(r) = 0 \quad (13)$$

corresponding to the saddle point and critical conditions respectively or from imposing the saddle point condition and requiring that the composite field α is massless

$$\Gamma(0) = 0 \quad (14)$$

The details of the calculations as well as the extension to an arbitrary potential can be found in Ref.⁷⁾.

It is worth noticing that, although the functional integral based on the action (1) makes sense only for $g > 0$, it is possible to extend its meaning in the framework of the large N expansion by taking care of the extra jacobian factor obtained in going to polar coordinates. This is the factor appearing on the r.h.s of the big bracket in eq. (12).

By calling ϵ a small quantity, that measures the distance from the critical coupling constant, and assuming naive scaling, that amounts to neglect ultraviolet divergences in the loops involving the composite field α , we get the following critical behaviour

$$F_\ell \sim \epsilon^{(1+1/k)(1-\ell)+\ell(k-1)D/2k} \\ F_1 \sim \log \epsilon \quad \text{if} \quad \ell = 1, D = 0 \quad (15)$$

where

$$F = -\log Z \quad (16)$$

has the following expansion for large N

$$F = NF_0 + F_1 + \frac{1}{N}F_2 + \dots = \sum_{\ell=0}^{\infty} \frac{F_\ell}{N^{\ell-1}} \quad (17)$$

As a consequence of the behaviour given in eq. (15) a double scaling behaviour can be defined by sending $N \rightarrow \infty$ and $\epsilon \rightarrow 0$ and by keeping the scaling variable

$$t = N\epsilon^{(k+1)/k-D(k-1)/(2k)} \quad (18)$$

fixed.

For $k = 2$ the behaviour in eq. (15) agrees with the expression obtained in a branched polymer model⁸⁾. This is not a coincidence³⁾, but it is connected to the fact that, unless in the matrix models where the large N expansion is a topological expansion, in vector models is a loop expansion. This means that the diagrams contributing at a certain order ℓ are all those containing ℓ loops of the composite field α . In terms of the diagrams with the original field $\vec{n}$ they have the structure of branched polymers.

The result given in eq. (15) is certainly correct for the cases $D = 0, 1$ where there are no ultraviolet divergences. For the higher dimensional cases a more careful analysis is needed.

In fact the naive behaviour determined by the infrared behaviour of the massless composite field gets spoiled by the renormalization procedure necessary to eliminate the ultraviolet divergences. For instance for $D = 2$ it can be seen that one gets an extra factor $\log \epsilon$ in each diagram in which there is a tadpole where the composite field is circulating. This implies that for $D = 2$ the behaviour for small ϵ given in eq. (15) gets an additional factor $(\log \epsilon)^n$ where n is the number of tadpoles appearing in the diagram. At a given number ℓ of loops the diagrams with ℓ tadpoles will be dominating and their behaviour will be given by eq.(15) times an extra factor $(\log \epsilon)^\ell$. Similar violations to the naive scaling behaviour appear also for higher values of D .

References

1. Brézin, E. and Kazakov, V.A., Phys. Lett. **236B**, 144 (1990).
Douglas, M.R. and Shenker, S., Nucl. Phys. **B355**, 635 (1990).
Gross, D.J. and Migdal, A.A., Phys. Rev. Lett. **64**, 127 (1990).
2. Ambjørn, J., Durhuus, B. and Frölich, J., Nucl. Phys. **B257**, 433 (1985).
David, F., Nucl. Phys. **B257**, 45 (1985).
Kazakov, V.A., Phys. Lett. **B150**, 282 (1985).
3. Nishigaki, S. and Yoneya, T., Nucl. Phys. **B348**, 787 (1991).
4. Anderson, A., Myers, R. C. and Periwai, V., Phys. Lett. **254B**, 89 (1991).
5. Di Vecchia, P., Kato, M. and Ohta, N., Nucl. Phys. **357**, 495 (1991).
6. Zinn-Justin, J., Phys. Lett. **257B**, 335 (1991).
7. Di Vecchia, P., Kato, M. and Ohta, N., Nordita preprint 91/22 P (1991). To be published in the Int. Jour. of Mod. Phys..
8. Ambjørn, J., Durhuus, B. and Jónsson, T., Phys. Lett. **B244**, 403 (1990).

$O(N)$ VECTOR FIELD THEORIES IN THE DOUBLE SCALING LIMIT

J. Zinn-Justin

Service de Physique Théorique, CE-Saclay

F-91191 Gif-sur-Yvette Cedex, FRANCE

zinn@poseidon.cea.fr

Abstract. It has been recently recognized that $O(N)$ invariant vector models have, in zero dimension, non-trivial scaling large N limits, a property they share with matrix models of 2D quantum gravity. We present here the extension of such analysis to field theory vector models in d dimensions and discuss the nature and form of the critical behaviour. We show that there exists, at least in some perturbative sense, a double scaling limit where a boundstate of the N -component vector ϕ associated with the ϕ^2 composite operator becomes massless, while the fundamental field ϕ itself remains massive. The limiting model can be described by an effective local interaction for the corresponding $O(N)$ invariant field. It is hoped that this investigation will lead to some new ideas about the behaviour of 2D quantum gravity coupled to $d > 1$ matter fields.

1. Introduction

It has been noticed in recent papers ^{[1][2][3]} that vector models have, in zero dimension, in the large N limit, critical points and scaling behaviours reminiscent

¹Laboratoire de la Direction des Sciences de la Matière du Commissariat à l'Energie Atomique

of those of matrix models relevant to 2D quantum gravity^{[4] [5] [6]}. These critical points seem to have a natural physical interpretation in the theory of branched polymers or filamentary surfaces^{[7], [1]}. We report here the extension of the analysis to vector models in quantum mechanics and d dimensional field theories^[8].

As we shall show, these peculiar critical points of vector models have the following origin: In the large N limit, vector models are reduced to perturbation theory around minima of an effective potential resulting from the competition between an attractive potential and a measure term which suppresses contributions of small fields. A singularity occurs when the minimum of the effective potential disappears. In matrix models also there is a competition between the potential and a measure term which produces a repulsion between eigenvalues, although the detailed mechanism is slightly different. In zero and one dimensions the repulsion is equivalent to the Pauli principle for a system of N independent fermions. A critical point is found when the Fermi level reaches an extremum of the potential.

Vector and matrix models share also other properties: Perturbation theory is always divergent, though like $k!$ in vector models instead of $2k!$ in matrix models^[9]. Moreover half of the vector models (like for matrix models) are ill-defined beyond perturbation theory because perturbation theory is non-Borel summable and the corresponding potential is not bounded from below.

2. Vector Models in Zero Dimension

The partition function of a field theory model in zero dimension is of course just given by an ordinary integral yielding sum rules for the weight factors of Feynman diagrams. We consider here the integral over a N -component vector $\mathbf{x}$ of an $O(N)$ invariant function of the form

$$Z = \int d\mathbf{x} e^{-NV(\mathbf{x}^2)}, \quad (2.1)$$

where $V(\rho)$ is a polynomial. Due to the $O(N)$ invariance of the integrand, we can integrate over angular variables and remain with an integral over $\rho = x^2$

$$Z \propto \int_0^\infty \frac{d\rho}{\rho} e^{-N[V(\rho) - \frac{1}{2} \ln \rho]}. \quad (2.2)$$

In the large N limit the integral can be calculated by steepest descent. The saddle point ρ_s is given by:

$$2V'(\rho_s)\rho_s = 1. \quad (2.3)$$

Critical points are points where the derivative of $V'(\rho)$ at ρ_s vanishes so that the result, at leading order, is no longer given by the gaussian approximation. Let us assume more generally that $p - 1$ successive derivatives vanish. The critical potential $V = V_c$ then satisfies

$$2V'_c(\rho)\rho - 1 = O((\rho_s - \rho)^p), \Rightarrow V_c(\rho) - \frac{1}{2} \ln \rho = \text{const.} + O((\rho_s - \rho)^{p+1}).$$

We see that the values of $\rho - \rho_s$ contributing to the integral are of order $N^{1/(p+1)}$. Let us add to the critical function V_c the set of relevant perturbations (in the sense of critical phenomena)

$$V(\rho) = V_c(\rho) + \sum_{q=1}^{p-1} v_q (\rho_s - \rho)^q,$$

(the term $q = p$ can be eliminated by a shift of ρ). Rescaling $\rho_s - \rho$ into $(\rho_s - \rho)N^{-1/(p+1)}$ we see that the scaling region corresponds to take $v_q = O(N^{(q-p-1)/(p+1)})$. Setting then

$$\rho_s - \rho = zN^{-1/(p+1)}, \quad u_q = N^{(p+1-q)/(p+1)}v_q,$$

we find the scaling “free energy” $F = \ln Z$:

$$F(u_q) \sim \ln \left\{ \int dz \exp \left[-z^{p+1} + \sum_{q=1}^{p-1} u_q z^q \right] \right\}. \quad (2.4)$$

Note that the case p odd corresponds to convergent integrals while the case p even can be only reached by analytic continuation and yields complex functions. A similar situation is found in matrix models. This problem will arise again in quantum mechanics and field theory and thus an analytic continuation will be, when necessary, implicitly assumed.

3. Quantum Mechanics

We consider now the partition function Z

$$Z = \int [d\mathbf{x}(t)] e^{-S(\mathbf{x})}, \quad (3.1)$$

for the $O(N)$ invariant action

$$S = N \int dt \left[\frac{1}{2} (\dot{\mathbf{x}})^2 + \frac{1}{2} \mathbf{x}^2 + \frac{g}{4} (\mathbf{x}^2)^2 \right]. \quad (3.2)$$

At the end of the section we shall briefly examine the problem of the existence of a large N scaling limit, using radial coordinates, from the Schrödinger equation point of view (the only method available in the matrix case). However, because we want to extend the analysis to field theory, we use here a more general method. We start from the identity

$$\exp -\frac{Ng}{4} \int dt (\mathbf{x}^2)^2 \propto \int [d\lambda] \exp N \int dt \left[\frac{\lambda^2}{4g} - \frac{\lambda}{2} \mathbf{x}^2 \right]. \quad (3.3)$$

Introducing identity (3.3) into the path integral, we can integrate over $\mathbf{x}$ and find

$$Z \propto \int [d\lambda(t)] e^{-S_{\text{eff}}(\lambda)}, \quad (3.4)$$

with

$$S_{\text{eff}}(\lambda) = \frac{N}{2} \left[- \int dt \frac{\lambda^2(t)}{2g} + \text{tr} \ln (-d^2 + 1 + \lambda) \right]. \quad (3.5)$$

The dependence on N of the partition function is now explicit. In the large N limit the path integral can be calculated by steepest descent. We look for a saddle point $\lambda(t) = \lambda$ constant which is then given by the equation:

$$\frac{\lambda}{g} = \frac{1}{2\pi} \int \frac{d\omega}{\omega^2 + 1 + \lambda} = \frac{1}{2\sqrt{1+\lambda}}. \quad (3.6)$$

A critical point is found when the equation has a double root. The condition is

$$1 + \frac{g}{4(1+\lambda)^{3/2}} = 0. \quad (3.7)$$

This equation implies that g is negative, and thus the potential is not bounded from below. The analysis which follows has thus no meaning beyond perturbation theory. With our normalizations the explicit values are

$$g_c = -4/3^{3/2} < 0, \quad \lambda_c = -2/3. \quad (3.8)$$

The $1/N$ -expansion. To generate the $1/N$ expansion we have first to calculate the λ -propagator. It is convenient to set:

$$\mu^2 = 1 + \lambda \Rightarrow \mu_c^2 = 1/3.$$

The Fourier transform $\Delta_\lambda(\omega)$ of the λ -propagator is then:

$$\Delta_\lambda^{-1}(\omega) = -\frac{N}{2} \left[\frac{1}{g} + \frac{1}{2\pi} \int \frac{d\omega'}{(\omega'^2 + \mu^2) [(\omega - \omega')^2 + \mu^2]} \right]. \quad (3.9)$$

The criticality condition implies that the inverse propagator vanishes at $\omega = 0$. Note that since μ_c does not vanish the inverse-propagator remains a regular function of ω^2 , which behaves like ω^2 for ω small. In higher dimensions we would expect λ to become a massless field in the critical limit. However, as we know, this is impossible in one dimension. The IR divergences are too strong. To understand the physics of the problem we perform a local (small ω in Fourier transform) and small $\lambda - \lambda_c$ expansion of the action. We then rescale time and λ :

$$t \mapsto tN^{1/5} \iff \omega \mapsto \omega N^{-1/5} \quad \lambda - \lambda_c \mapsto \tilde{\lambda} N^{-2/5}, \quad (3.10)$$

to render N independent the coefficients of $\tilde{\lambda}^3$ and $(\dot{\tilde{\lambda}})^2$ in the action. We note that terms with higher derivatives or higher powers of $\tilde{\lambda}$ are then suppressed for N large. Therefore at leading order for N large (and after some additional finite rescaling) the action takes the form:

$$S(\tilde{\lambda}) \sim \int dt \left[\frac{1}{2} \left(\frac{d\tilde{\lambda}}{dt} \right)^2 + \frac{\tilde{\lambda}^3}{3} \right]. \quad (3.11)$$

Of course the action is not bounded from below and the path integral can only be defined by analytic continuation. To action (3.11) corresponds the hamiltonian H :

$$H = N^{-1/5} \left(-\frac{1}{2} (d/d\tilde{\lambda})^2 + \frac{1}{3} \tilde{\lambda}^3 \right).$$

The scaling region. For g close to g_c the most relevant new interaction term is the term linear in $\tilde{\lambda}$. After the rescaling (3.10) it gives an additional contribution to the action proportional to $(g - g_c)N^{4/5}\tilde{\lambda}$. The scaling region is thus defined by letting N go to infinity and $g - g_c$ to zero at

$$(g - g_c)N^{4/5} = u$$

fixed. The hamiltonian relevant to the scaling limit is then:

$$H = N^{-1/5} \left[-\frac{1}{2} \left(\frac{d}{d\tilde{\lambda}} \right)^2 + u\tilde{\lambda} + \frac{\tilde{\lambda}^3}{3} \right]. \quad (3.12)$$

We have found a scaling regime analogous to the one observed in $d < 1$ matrix models. In the $d = 1$ matrix model, instead, the situation is more complicated [10] because logarithmic deviations from a simple scaling law are found, a situation we shall meet in the $d = 2$ vector model.

Hamiltonian formalism. For N large the zero angular momentum hamiltonian H_0 can be written:

$$H_0 = -\frac{1}{2N} \left(\frac{d}{dr} \right)^2 + NW(r), \quad (3.13)$$

where $r = |\mathbf{x}|$ and

$$W(r) = \frac{1}{8r^2} + \frac{r^2}{2} + g\frac{r^4}{4},$$

For N large we can expand perturbation theory around the classical minimum of the potential. This minimum becomes a relative minimum when g is negative. It disappears at a value of g where also the second derivative of $W(r)$ vanishes. A short calculation yields:

$$r^2 = 3^{1/2}/2, \quad g = -4/3^{3/2}.$$

We recognize the critical value (3.8) of g found above.

Generalization. It is easy to extend previous analysis to more general $O(N)$ invariant interactions. The large N expansion can be obtained from standard techniques (see for example [11]). Let us consider the general action:

$$S = N \int dt \left[\frac{1}{2} (\dot{\mathbf{x}})^2 + V(\mathbf{x}^2) \right]. \quad (3.14)$$

We introduce a Lagrange multiplier λ to impose the constraint $\mu = \mathbf{x}^2$. The action then takes the form

$$S(\mathbf{x}, \lambda, \mu) = N \int dt \left[\frac{1}{2} (\dot{\mathbf{x}})^2 + V(\mu) + \frac{1}{2} \lambda (\mathbf{x}^2 - \mu) \right]. \quad (3.15)$$

We integrate over $\mathbf{x}$ to obtain:

$$S(\lambda, \mu) = N \left\{ \int dt \left[V(\mu) - \frac{1}{2} \lambda \mu \right] + \frac{1}{2} \text{tr} \ln (-d^2 + \lambda) \right\}. \quad (3.16)$$

In the large N limit the path integral can again be calculated by steepest descent. We look for two constants λ, μ solutions of

$$V'(\mu) - \frac{1}{2} \lambda = 0, \quad \mu = \frac{1}{2\pi} \int \frac{d\omega}{\omega^2 + \lambda} = \frac{1}{2\sqrt{\lambda}}. \quad (3.17)$$

When the determinant of partial derivatives of this system vanishes we find a critical point. A short calculation yields the set of equations:

$$8\mu^2 V'(\mu) = 1, \quad 4\mu^3 V''(\mu) = -1.$$

The matrix of second derivatives of the action is the mass matrix in field theory language. Diagonalizing this matrix we understand from the previous analysis that this implies that the propagator of a combination $\tilde{\lambda}$ of λ and μ has a pole at $\omega = 0$. The field corresponding to the second eigenvector remains massive and can be integrated out. We are back to the problem examined above. The main difference with the case of the quartic interaction is that we can now adjust the original potential V in such a way that the coefficient of the leading $\tilde{\lambda}^3$ interaction vanishes, or more generally all interactions up to the $\tilde{\lambda}^p$ vanish. We then perform a local expansion of the action and rescale time and $\tilde{\lambda}$ to render the coefficients of $(\tilde{\lambda})^2$ and $\tilde{\lambda}^{p+1}$ N independent:

$$t \mapsto tN^{(p-1)/(p+3)} \iff \omega \mapsto \omega N^{-(p-1)/(p+3)} \quad \lambda - \lambda_c \mapsto \tilde{\lambda}N^{-2/(p+3)}. \quad (3.18)$$

It is easy to verify that interactions containing derivatives or higher powers of $\tilde{\lambda}$ are then suppressed for N large.

To describe the scaling region we can add to the critical potential a set of relevant terms characterized by parameters v_q . The scaling limit is then obtained by keeping the products u_q ,

$$u_q = N^{(2p+2-2q)/(p+3)} v_q, \quad q = 1, \dots, p-1,$$

fixed and the corresponding scaling hamiltonian is:

$$H = N^{-(p-1)/(p+3)} \left[-\frac{1}{2} \left(\frac{d}{d\tilde{\lambda}} \right)^2 + \frac{\tilde{\lambda}^{p+1}}{p+1} + \sum_{q=1}^{p-1} u_q \frac{\tilde{\lambda}^q}{q} \right]. \quad (3.19)$$

4. Field Theory

We come now to the most interesting case, field theory in $d > 1$ dimensions. From the algebraic point of view the analysis is very similar. However we expect one essential difference with the case of quantum mechanics $d = 1$. We have seen that in the critical limit, at leading order in $1/N$, one boundstate associated with the composite ϕ^2 field becomes massless (the ϕ -field itself remaining non-critical). This is the physics of the phase transition in an Ising-like system. While the phase transition is impossible for $d = 1$, it is possible in higher dimensions.

We discuss this problem first in the special case of the $(\phi^2)^2$ interaction. We consider the partition function:

$$Z = \int [d\phi(x)] e^{-S(\phi)}, \quad (4.1)$$

where $S(\phi)$ is a $O(N)$ symmetric action:

$$S(\phi) = N \int \left\{ \frac{1}{2} [\partial_\mu \phi(x)]^2 + \frac{1}{2} r \phi^2(x) + \frac{g}{4} [\phi^2(x)]^2 \right\} d^d x. \quad (4.2)$$

A cut-off of order 1, consistent with the symmetry, is now implied. For example we can assume that the inverse propagator has higher order derivative terms and we have explicitly written in action (4.2) only the two first terms in a local expansion (in Fourier space small momentum expansion). In particular, the parameter r is then the value of the inverse propagator at zero momentum.

We use again the identity

$$\exp -\frac{Ng}{4} \int d^d x (\phi^2)^2 \propto \int [d\lambda] \exp N \int d^d x \left[\frac{\lambda^2}{4g} - \frac{\lambda}{2} \phi^2 \right]. \quad (4.3)$$

Introducing this identity into (4.1) and performing the gaussian integration over ϕ we find

$$Z = \int [d\lambda(x)] e^{-S_{\text{eff}}(\lambda)}, \quad (4.4)$$

with:

$$S_{\text{eff}}(\lambda) = \frac{N}{2} \left[- \int \frac{\lambda^2(x)}{2g} d^d x + \text{tr} \ln(-\Delta + r + \lambda(x)) \right]. \quad (4.5)$$

In the large N limit the functional integral can be calculated by steepest descent.

The large N limit. We look for a uniform saddle point $\lambda(x) = \lambda$. The saddle point equation is:

$$-\frac{\lambda}{g} + \frac{1}{(2\pi)^d} \int \frac{d^d p}{p^2 + r + \lambda} = 0. \quad (4.6)$$

Let us introduce the parameter m :

$$m^2 = r + \lambda, \quad (4.7)$$

which is, in the large N limit, the mass of field ϕ . In terms of m the equation becomes:

$$\frac{(m^2 - r)}{g} - \frac{1}{(2\pi)^d} \int \frac{d^d p}{p^2 + m^2} = 0. \quad (4.8)$$

Expressing that the derivative with respect to λ or m^2 vanishes, we obtain an equation for the critical point:

$$\frac{1}{g} + \frac{1}{(2\pi)^d} \int \frac{d^d p}{(p^2 + m^2)^2} = 0. \quad (4.9)$$

These equations define, at r fixed, critical values of g and λ . The critical value g_c of g is again negative. Criticality equation (4.9) implies that the λ -propagator has a pole at zero momentum. The λ -field becomes massless while the ϕ -field remains massive.

5. The Scaling Limit in Field Theory

We have studied the large N limit. We now look for a scaling limit. Since the λ field is a one-component field it can remain critical for $d > 1$ even in presence of interactions. Still perturbation theory is IR divergent for dimensions $d \leq 6$. Let us therefore immediately add a relevant perturbation proportional to $g - g_c$ which will provide the theory with an IR cut-off. Let us then look for the most IR divergent terms in perturbation theory. We face a standard problem in the theory of critical phenomena [11]: The bare mass squared which is proportional to $\sqrt{g_c - g}$ plays exactly the role of a deviation from the critical temperature. The effective action for the λ -field is non-local and contains arbitrary powers of the field. However, because the ϕ -field is not critical we can again make a local expansion. Standard arguments of the theory of critical phenomena tell us that the most IR divergent terms come from interactions without derivatives and with the lowest power of the field. Here the leading interaction is proportional to λ^3 . To characterize the IR divergences of the perturbative expansion in powers in $1/N$ let us again rescale distances and field $\lambda \rightarrow \lambda_c$:

$$\lambda - \lambda_c \propto \tilde{\lambda} \Lambda^{(2-d)/2} N^{-1/2}, \quad x \mapsto \Lambda x, \quad (5.1)$$

where Λ will play the role of a cut-off. The effective action at leading order, after some additional finite renormalizations, is

$$S_{\text{eff}}(\tilde{\lambda}) = \int d^d x \left[\frac{1}{2} \left(\partial_\mu \tilde{\lambda} \right)^2 + v \Lambda^{(d+2)/2} \sqrt{N} \tilde{\lambda} + \frac{\Lambda^{(6-d)/2}}{3\sqrt{N}} \tilde{\lambda}^3 \right],$$

where $v \propto g - g_c$ and a cut-off Λ is now implied.

We first consider dimensions $d < 6$. The $\tilde{\lambda}^3$ field theory is then super-renormalizable and we fix the coefficient of $\tilde{\lambda}^3$:

$$\Lambda^{(6-d)/2} / \sqrt{N} = g_3.$$

Therefore the cut-off grows with N like $N^{1/(6-d)}$. However, unlike the case of quantum mechanics, we cannot also keep the quantity $v \Lambda^{(d+2)/2} \sqrt{N}$ fixed because the field theory has UV divergences when the cut-off Λ becomes large.

Dimensions $d < 4$. It is convenient to first examine dimensions $d < 4$ because then only the field average is divergent. Thus we have to introduce a counterterm which renders $\langle \tilde{\lambda}(x) \rangle$ finite. The renormalized action is

$$S_{\text{eff}}(\tilde{\lambda}_r) = \int d^d x \left[\frac{1}{2} \left(\partial_\mu \tilde{\lambda}_r \right)^2 + \frac{1}{2} \mu^2 \tilde{\lambda}_r^2 + \frac{g_3}{3} \tilde{\lambda}_r^3 - c_1(\Lambda) \tilde{\lambda}_r \right],$$

in which $\tilde{\lambda}_r$ is the renormalized field and μ is a renormalized mass parameter.

To recover the original action we must eliminate the term quadratic in the field and thus shift $\tilde{\lambda}_r$ by a quantity $\bar{\lambda}$:

$$\bar{\lambda} = -\mu^2/2g_3.$$

Identifying then the coefficients of the linear term we find:

$$v\Lambda^{(d+2)/2}\sqrt{N} = -c_1(\Lambda) - \frac{\mu^4}{4g_3}.$$

For $d = 2$ only the one-loop diagram is divergent and one finds

$$c_1(\Lambda) = \frac{g_3}{2\pi} \ln(\Lambda/\mu).$$

Therefore to obtain a non-trivial scaling limit we have to choose*:

$$v = -\frac{1}{N} \left[\frac{1}{8\pi} \ln(Ng_3^2/\mu^4) + \frac{\mu^4}{4g_3^2} \right]. \quad (5.2)$$

Only the $\ln N/N$ term is universal, the $1/N$ term is regularization and renormalization dependent.

For $d = 3$ the two-loop diagram is also divergent. Still for Λ large the leading contribution is still given by the one-loop diagram, thus $c_1(\Lambda) \propto \Lambda$ and $v = O(1/N)$. We have thus shown that for $d < 4$ a scaling limit exists which leads to a renormalized λ^3 field theory. The relation between $g - g_c$ and N , however, has itself no longer a simple power law form.

Note finally the similarity between the results for the vector model at $d = 2$ and the matrix model at $d = 1$ ^[10].

Dimensions $4 \leq d < 6$. For $4 \leq d < 6$ the situation is slightly more complicated because two counterterms are needed, renormalizing $\langle \tilde{\lambda} \rangle$ and the coefficient of $\tilde{\lambda}^2$. The quantity $v \propto (g - g_c)$ becomes a even more complicated function of N . However, at leading order for N large, v still behaves as $1/N$, while naive scaling would have predicted a scaling variable $Nv^{(6-d)/4}$.

Higher dimensions. For $d = 6$ the theory is just renormalizable. As we know renormalization group arguments are required to discuss the effect of IR divergences. Since the $\tilde{\lambda}^3$ field theory is not defined this discussion is meaningless except if we assume that we have adjusted the original coupling constant in such a way that the coefficient of $\tilde{\lambda}^3$ is pure imaginary, as in the case of the Lee-Yang edge singularity. Then the theory is IR free in six dimensions, and in the absence of any other non-trivial IR fixed point, it is impossible to define a scaling limit. The same applies to higher dimensions where the theory is non-renormalizable, and therefore the powers of $1/N$ cannot be compensated by IR divergences.

* Note that there is a misprint in the corresponding equation in ref. [8].

6. More General Interactions

More general interactions can be dealt with by the method explained in the case of quantum mechanics. A cut-off has, however, to be introduced to make the vector models meaningful in all dimensions. The results are very similar. Because only one mode is critical the main effect is to introduce additional parameters in the effective interaction of the critical mode in such a way that the most IR relevant interactions can be cancelled. In the language of critical phenomena we reach multicritical points. We then generate the renormalized λ^{p+1} interactions provided we again choose the parameters of the potential, as functions of N , such that they cancel the UV divergences of perturbation theory. Let us briefly examine the case of the $\tilde{\lambda}^4$ field theory. There are two relevant terms in the IR region proportional to $\tilde{\lambda}$ and $\tilde{\lambda}^2$. Let us call v_1, v_2 the corresponding amplitudes. After rescaling (5.1) the action reads

$$S_{\text{eff}}(\tilde{\lambda}) = \int d^d x \left[\frac{1}{2} \left(\partial_\mu \tilde{\lambda} \right)^2 + v_1 \Lambda^{(d+2)/2} \tilde{\lambda} + \frac{1}{2} v_2 \Lambda^2 \tilde{\lambda}^2 + \frac{\Lambda^{4-d}}{4!N} \tilde{\lambda}^4 \right].$$

For $d < 4$ only a mass counterterm is needed. We first set:

$$\frac{\Lambda^{4-d}}{N} = g_4, \quad v_1 \Lambda^{(d+2)/2} = u_1 \Rightarrow v_1 = O\left(N^{-(d+2)/2(4-d)}\right),$$

and

$$v_2 \Lambda^2 = \mu^2 - c_2(\Lambda),$$

where $c_2(\Lambda)$ is the mass renormalization constant. In two dimensions we find:

$$v_2 = \frac{1}{N g_4} \left(\mu^2 + \frac{g_4}{8\pi} \ln(N g_4 / \mu^2) \right).$$

In dimension $d = 3$ the relation becomes more complicate because several diagrams are primitively divergent. Note, however, that at leading order for N large $v_2 = O(1/N)$, while naive scaling would have predicted $v_2 = O(1/N^2)$.

In dimensions $d > 4$ perturbation theory is no longer IR divergent and therefore no scaling limit can be found. For $d = 4$ the IR divergences can be summed using renormalization group arguments. Since the $\tilde{\lambda}^4$ field theory is IR free no non-trivial scaling limit can be found either.

Finally for $d = 3$ only one other non-trivial critical theory exists which corresponds to a $\tilde{\lambda}^5$ interaction. Only in two dimensions can one construct an infinite sequence of different critical limits.

7. $O(N)$ Invariant Rectangular Matrix Models

It is possible to study by the same method a class of models slightly closer to the matrix models of 2D gravity. The fields are rectangular complex $n \times N$ matrices M and the action is

$$S(M) = N \int d^d x \left[\text{tr} \partial_\mu M \partial_\mu M^\dagger + r \text{tr} M M^\dagger + \frac{1}{2} g \text{tr} (M M^\dagger)^2 \right].$$

The large N limit is then taken at n fixed. The quartic interaction term can be generated by the usual trick, as the result of a gaussian integration over a $n \times n$ hermitian matrix λ . After integration over the matrix M we obtain the effective action:

$$S_{\text{eff}} = N \left[- \int d^d x \frac{\text{tr} \lambda^2}{2g} + \text{tr} \ln(-d^2 + r + \lambda) \right].$$

The saddle equation tell us to shift λ by a multiple of the identity. The critical point is reached when the mass of the λ -field vanishes. We are in a situation very similar to the previous case, except that the order parameter is now a $n \times n$ hermitian matrix and the model has a continuous $U(n)$ symmetry. For $d = 1$ a hamiltonian with a $\text{tr} \lambda^3$ interaction is generated. For $d = 2$ we expect a subtle difference with the previous examples because no phase transition is possible in a system with continuous symmetry.

For more general potentials we have to introduce two hermitian matrices. Formally we can then generate more general interactions of $\text{tr} \lambda^p$ type. Note, however, that for n large enough many of such matrix models have first order phase transitions even for p even.

8. Concluding Remarks and Prospects

We have shown on a few simple examples that vector models have, at least in low dimensions, non-trivial large N scaling limits. Obviously more general models can be analyzed by similar methods. Of particular interest is the generalization to models containing fermions (in particular supersymmetric models). One would like to find out whether the presence of fermions stabilizes some of the models which are unstable otherwise, a very important problem for the corresponding matrix models of 2D gravity.

References

- [1] Anderson A., Myers R.C. and Perival V., Phys. Lett. **B254**, 89 (1991).
- [2] Nishigaki S. and Yoneya T., Nucl. Phys. **B348**, 787 (1991).
- [3] Di Vecchia P., Kato M. and Ohta N., Double scaling limit in $O(N)$ vector models, Nordita preprint ISSN-0106-2646, Nucl. Phys. to be published.
- [4] David F., Nucl. Phys. **B257**[FS14], 45, 543 (1985);
Ambjørn J., Durhuus B. and Fröhlich J., Nucl. Phys. **B257**[FS14], 433 (1985);
Kazakov V.A., Phys. Lett. **B150**, 282 (1985).
- [5] Douglas M. and Shenker S., Nucl. Phys. **B335**, 635 (1990);
Brézin E. and Kazakov V., Phys. Lett. **B236**, 144 (1990);
Gross D. and Migdal A.A., Phys. Rev. Lett. **64**, 127 (1990); Nucl. Phys. **B340**, 333 (1990).
- [6] Douglas M. R., Phys. Lett. **B238**, 176 (1990);
David F., Mod. Phys. Lett. **A5**, 1019 (1990);
Ginsparg P., Goulian M., Plesser M.R. and Zinn-Justin J., Nucl. Phys. **B342**, 539 (1990);
Jevicki A. and Yoneya T., Mod. Phys. Lett. **A5**, 1615 (1990).
- [7] Ambjørn J., Durhuus B. and Jonsson T., Phys. Lett. **B244**, 403 (1990).
- [8] Zinn-Justin J., Phys. Lett. **B257**, 335 (1991).
- [9] Ginsparg P. and Zinn-Justin J., Phys. Lett. **B255**, 189 (1991);
Ginsparg P. and Zinn-Justin J., "Action principle and large order behavior of non-perturbative gravity", LA-UR-90-3687/SPhT/90-140 (1990), to appear in *Random surfaces and quantum gravity*, proceedings of 1990 Cargèse workshop, edited by O. Alvarez, E. Marinari, and P. Windey.
- [10] Ginsparg P. and Zinn-Justin J., Phys. Lett. **B240**, 333 (1990);
Brézin E., Kazakov V. A., and Zamolodchikov Al. B., Nucl. Phys. **B338**, 673 (1990);
Parisi G., Phys. Lett. **B238**, 209, 213 (1990); Europhys. Lett. **11**, 595 (1990);
Gross D. J. and Miljkovic N., Phys. Lett. **B238**, 217 (1990).
- [11] J. Zinn-Justin, *Quantum field theory and critical phenomena*, Oxford Univ. Press 1989.

This page is intentionally left blank

This page is intentionally left blank

This page is intentionally left blank

